Metacognition in face processing

Xingchen Zhou

Doctor of Philosophy

University of York
Psychology

March 2021

Abstract

This thesis brings together two streams of research that have traditionally proceeded independently: metacognition and face processing. In the first experimental chapter, I assessed participants' insight into their own face perception abilities and those of other people. I found classic Dunning–Kruger Effects in matching tasks for unfamiliar identity, familiar identity, gaze direction, and emotional expression. Low performers overestimated themselves, and high performers underestimated themselves. Interestingly, participants' self-estimates were more stable across tasks than their actual performance. In addition, peer estimates revealed a consistent egocentric bias. High performers attributed higher accuracy to other people than did low performers. In the second experimental chapter, I focus on the other-race effect (ORE), the phenomenon whereby own-race faces are better remembered than other-race faces. Despite the high profile of ORE in the literature, previous studies have not put the magnitude of ORE in context. In two face recognition experiments, I show that the familiarity effect was several times larger than the race effect in all performance measures. However, participants expected race to have a larger effect on others than it actually did. Face recognition accuracy depends much more on whether you know the person's face than whether you share the same race. In the final experimental chapter, I examined how much common ground exists in face-evoked thoughts, and how the observed overlap compares to viewers' expectations. I show that participants exhibited strong egocentric bias and false consensus effects, greatly overestimating the extent to which other people's thoughts resembled their own. The findings of this thesis have both theoretical and applied implications. Not only do they shed light on the range of ability in the general population, they also reveal a fundamental source of uncertainty in social interactions.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Rob Jenkins for his unwavering support and invaluable supervision during each stage of my PhD study. Rob continuously provided encouragement and was always willing and enthusiastic to assist in any way he could in all the time of my academic research and daily life. Second, I would thank for Professor Steven Tipper, Professor Katie Slocombe, Dr. Tom Hartley and all the members at the face lab for their insightful comments and suggestions, especially Professor Mike Burton's help on Chapter 3. I would also thank V. Sefa and N. Ajoku from the African Caribbean Society, University of York, for help recruiting participants and for advice on terminology. Many thanks to all participants that took part in my experiments and enabled this research to be possible. My appreciation also goes out to my family and friends for their encouragement and support all through my studies.

Author's Declaration

I, Xingchen Zhou, declare that this thesis is my own work carried out under normal terms of supervision. This work has not been previously presented for an award at this, or any other University. All quotations in this thesis have been distinguished by quotation marks and they have been attributed to the original source. All sources are acknowledged as References.

The empirical work presented in this thesis has been published or is currently under review in the following peer-reviewed journals:

Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. *Cognition*, *203*, 104345. doi:10.1016/j.cognition.2020.104345

X. Zhou: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Supervision; Validation; Visualization; Writing - original draft. R. Jenkins: Conceptualization; Methodology; Project administration; Resources; Supervision; Visualization; Writing - review & editing. All authors approved the final version of the paper.

Zhou, X., A.M. Burton & Jenkins, R. (2020). Familiarity matters more than race in face recognition memory. Manuscript submitted for publication. [Currently under review in *Perception*]

X. Zhou and R. Jenkins conceived the study and designed the experiments. X. Zhou collected the data and conducted all statistical analyses. X. Zhou and R. Jenkins coordinated visualizations. X. Zhou drafted the manuscript. X. Zhou, R. Jenkins, and A. M. Burton wrote and edited the manuscript. All authors approved the final version of the paper.

Zhou, X., A.M. Burton & Jenkins, R. (2020). Face-evoked thoughts. Manuscript submitted for publication. [Currently under review in *Cognition*]

X. Zhou: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Supervision; Validation; Visualization; Writing - original draft. R. Jenkins: Conceptualization; Methodology; Project administration; Resources; Supervision; Visualization; Writing - review & editing. All authors approved the final version of the paper.

Primary Supervisor Statement

I am listed as a co-author on the three empirical papers which make-up the main body of this thesis.

In each of the reported studies, the work is primarily that of Ms. Xingchen Zhou. For each paper, Xingchen compiled the relevant stimuli, completed all of the data collection and coded and analysed the summary data. Xingchen wrote the first draft of each paper.

Dr. Rob Jenkins

Secondary Supervisor Statement

I am listed as second co-author on the empirical paper which makes-up Chapter 3 of this thesis.

This work is primarily that of Ms. Xingchen Zhou. Xingchen compiled the relevant stimuli, completed all of the data collection and coded and analysed the summary data. Xingchen wrote the first draft of this paper.

Prof. Mike Burton

Chapter 1 General Introduction

This introduction is broadly divided into two general themes. The first general theme is metacognition, including its definition, importance, measurements and two well-established phenomena—the Dunning–Kruger effect (concerning insight into one's own thinking) and egocentric bias (concerning insight into other people's thinking). The second broad theme is face perception, including its definition, importance, measurements and different aspects of face processing that are potentially connected to metacognition. The chapter ends with an overview of the experimental work contained in this thesis exploring metacognition and face perception.

1.1 What is metacognition

In 1995, a peculiar bank robbery was committed in Pittsburgh. The culprit, McArthur Wheeler, robbed two banks in broad daylight without any visible attempt at disguise. He could not understand how he had been arrested so quickly, because he had covered his face with lemon juice, believing that this precaution would make it difficult for the police to recognise him from the videotape (Fuocco, 1996). The Wheeler case vividly illustrates the fact that people are largely unaware of their own incompetence (Kruger & Dunning, 1999), and tend to surmise others' behavior according to their own.

1.1.1 Definition

Negotiating everyday life requires accurate and appropriate insight into one's own intellectual and social limitations. That leads to the focus of this project, which is "metacognition", generally defined as “cognition about cognition” or “thinking about thinking”. Metacognition has been defined in various ways (Kuhn, 1983), comprising ‘executive processes’ (Brown, 1977), ‘cognitive monitoring’ (Flavell, 1979) and ‘self-communication’ (Meichenbaum & Asarnow, 1979). The formal model of metacognition that has gathered a great deal of support and widely cited in recent years is the distinguishment between primary and secondary cognition (Nelson & Narens, 1990). Primary cognition is often called “object level” cognition, referring to the mental processes involved in gaining knowledge and comprehension. At the higher level is metacognition. A typical example to distinguish object level and the meta level is the well-known tip-of-the-tongue situation (e.g., You can’t recall a person’s name but you are sure that you know the person and the name), which may happen more frequently as people grow older (Schwartz, 2002). Metacognitive monitoring and metacognitive control are the two facets of metacognition that have been investigated most extensively, according to the framework of metacognition formalized by Nelson and Narens (1990). The definition and examples of the key concepts are presented in Table 1.1, (John & Janet, 2008). Metacognitive monitoring occurs when the object level (e.g., in this case, your thought is that “I know this person and the name”) is being assessed by the meta level (“I’m sure I know the person’s name”). Metacognitive control occurs when the meta level influences control over the object level (e.g., when the awkward tip-of-the-togue situation makes you spend more time to match the face and the name). Therefore, metacognitive control serves as a further process based on the results of metacognitive monitoring. Given the current work is the extension of metacognition research in a new field, the current work only considered the first process which is the metacognitive monitoring at this stage.

Table 1.1 Definitions of Important Concepts Relevant to Metacognition

Concept	Definition	Examples
Cognition	Symbolic mental activities and mental representations	thinking, learning, memory, reasoning, problem-solving
Metacognition	Cognition about cognition	See examples in text
Metacognitive monitoring	Assessing the current state of a cognitive activity	Judging whether you are approaching the correct solution to a problem Assessing how well you understand what you are reading
Metacognitive control	Regulating some aspect of a cognitive activity	Deciding to use a new tactic to solve a difficult problem Deciding to spend more time trying to remember the answer to a trivia question

Metacognitive thoughts vary in a number of interesting dimensions, such as valence, number, target, origin, evaluation, and confidence (Petty et al., 2007). This has to some extent explained why metacognition is such a complex concept to define. There is a long-standing historical controversy in metacognition definition between researchers who stressed that metacognition only contains the thinking about one's own cognition (e.g., Flavell, 1979; Kuhn, 2000; Martinez, 2006) versus those who conceived metacognition more broadly as thinking about their own and other people's cognition (e.g., Jost, Kruglanski, & Nelson, 1998; Wright, 2002). The prediction of other people's performance—peer judgement—was also included as a part of metacognition in Tauber, Dunlosky, Rawson, Rhodes, and Sitzman (2013). Thus, this faculty can be further subcategorized into beliefs about intra-individual differences and inter-individual differences. In other words, metacognition refers to people's insight into themselves and others (any thought about a thought). However, the ability to understand the thoughts of others and to use this information to predict the behavior of others is known as mentalizing (also known as Theory of Mind; Frith & Frith, 1999; Frith, 2012). Theory of mind has been widely used in developmental research, because it focuses on the ability to distinguish between people’s own beliefs and those of others (Keysar, Lin, & Barr, 2003). Importantly, self-assessment can often end up in a chicken-and-egg relationship with the assessment of others. Knowledge of oneself may serve as the first step towards understanding others (Couchman, Coutinho, Beran, & David Smith, 2009). On the other hand, we can better know ourselves by understanding others (Tokuhama-Espinosa, 2014). In addition, brain studies also found that these two modes of assessment involve similar neural networks (Legrand & Ruby, 2009). In the present thesis, the definition of metacognition will be considered in a general way as the estimation of both one’s own performance and that of others.

1.1.2 Theoretical and practical importance

The presence and consequences of metacognition have significant implications for the theory of understanding human consciousness. Metacognition is a higher level of cognition that does not show superficial behavior but rather the mental states behind the behavior (Nelson, 1996; Koriat, 2007). Understanding this kind of reflective knowledge is also important for a complete understanding of human behavior. According to Petty, et al. (2007), metacognitive performance can also influence cognitive behavior. Take the behavior of pursuing goals as an example, which is an important step associated with desirable outcomes. As noted by Achtziger, et al. (2011), metacognition influence future goal setting and goal striving. That is, the reflection and evaluation is relevant in determining a new goal and how much effort is required to reach the goal. Besides helping understand ourselves, according to the other side of metacognition, it also contributes to the complete understanding of others. Couchman et al. (2009) argued that metacognition is a prior step towards understanding the thoughts of others and taking their perspective. The assessment of other people's performance is a primary basis of social interaction (Fussell & Krauss, 1989; Hardin & Higgins, 1996). Metacognition is not only a key component of humans' reflective mind, cognitive functioning and interpersonal communication in theory, it has also broadened the understanding of learning by educational psychologists. Two people could have the same performance at the cognition level, but their performance at the meta level could be quite different. For example, suppose that two excellent students have already mastered an area of knowledge and both chose option A (the right answer) in a test question. One student had great self-assessment and didn't doubt herself or himself. The other student did not have such well-calibrated judgment of her or his own performance. After some hesitation, the correct option A was unfortunately changed to the wrong option B. Thus, the study of metacognition can help to understand what differentiates successful students from their less successful peers in learning.

Just as one's knowledge of a goal state is likely to guide the specific strategy you will use to achieve your goal, the study of metacognition not only enriches theorietical understanding of human behavior, but also directly influences human behavior. Therefore, the topic of metacognition holds an important place from a practical perspective, especially in learning and social interactions. For example, research shows that metacognition is significantly correlated to academic monitoring (Sperling, et al. , 2004) and is also related to predictions of subsequent learning (Tobias & Everson, 1996). That is, once people perceive that their degree of learning meets the required standard, they may not continue learn the current item. Those who have poor self-assessment never knew that they haven't grasped the knowledge well, so they arrange little time in learning due to their overconfidence, leading to negative impacts on their study outcomes. Therefore, since the metacognitive ability can promote academic learning, educational psychologists are keen to explore strategies for improving metacognition performance. Many teachers who realize the importance of metacognition skills also explicitly teach learners metacognition strategies and cultivate them as a mental habit in order to improve in attainment (Martinez, 2006). Meanwhile, better assessment of other people's behavior can help people understand others better in interpersonal communication, which may also help to reduce contradiction and conflict, and promote cooperation (Frith, 2012). In the field of social psychology, stereotype has become a major concern for social scientists. According to Banaji and Dasgupta (1998), metacognitive improvement can be an entry point to dispel stereotypes. For instance, if people are aware of the impact of their stereotypic biases, they could adjust the effect of the bias, or perhaps the bias would not exist in the first place. This is not an isolated case. For example, a skilled driver may think it easy to dodge a car rushing towards him, and may think it easy for the driver of the other car too. Yet the driver in the other car may be a novice. That kind of unawareness may increase the risk of a traffic accident. Hence, metacognition, the knowledge of knowledge, is as important as knowledge *per se*.

1.2 Metacognition measurements

1.2.1 Types of measurements

The measurements available to researchers in the previous metacognition studies vary in different ways. Here, I review the following typical behavioral methods based on different distinctions (see table 1.2, note that there are still more measurements based on other distinctions or other domains). The first distinction is timing, including whether measurements of metacognitive performance occur prior to the whole cognitive task (prospective, e.g., Schnitzspahn et al., 2011), follow each trial (concurrent, e.g., Mclntosh et al., 2019) or after task completion (retrospective, e.g., Mazancieux et al. 2018).

Table 1.2 Summary of metacognitive measurements classified by timing and domain.

	domain at cognitive level			
timing	memory	decision-making	perception	learning
prospective	judgement of learning	performance estimate	performance estimate	performance estimate
concurrent	feeling of knowing	confidence	confidence	think-aloud
retrospective	confidence	wager	a separate scale	assess action

Metacognition has been widely explored in various domains, but measurement in one domain is difficult to transfer directly to another domain. Thus, another distinction of metacognition measurement is domain. In the memory domain, judgements of learning (JOL) mainly occur in the learning phase testing self-assessment of how successfully participants will recall a particular item in a subsequent test phase (Hu et al., 2016). Conversely, feeling of knowing (FOK) mainly occurs after the test phase when participants fail to recall an item, or to predict the probability that they might be able to recognize the answer from a list of alternatives (Hart, 1965). FOK is closely related to the tip-of-the-tongue phenomenon (Brown, 1991).

Metacognitive confidence refers to the estimation of the possibility that the answer could be correct usually along with clear criteria for accuracy. Asking for confidence-in-accuracy has been widely used in a variety of domains, including memory (Brewer & Sampaio, 2012), decision-making (Yeung & Summerfield, 2012), and visual perception (Vickers, 2014). For example, Rausch and colleagues (2015) asked observers to rate their degree of confidence in a perceptual judgement using a two-alternative forced-choice (“unconfident” and “confident”). Researchers have also used other scales such as the four-point confidence scale in Peirce & Jastrow (1884). Estimating specific performance is another widely-used measure. For example, asking participants to guess how well other people would perform in a face-matching task after making the judgement by themselves (Ritchie et al., 2015). Decision-making researchers also used wagers as an intuitive measure of retrospective confidence, as introduced by Kunimoto et al., (2001). In the standard simple post-decision wagering, participants are asked to bet on whether their answer is correct or not. They will keep the wager amount if their decision is correct. Otherwise, they will lose it. Their confidence is reflected by the size of chosen gamble (e.g., Persaud, et al., 2007).

In addition to the respective measurement methods in each field, researchers have also developed metacognitive scales to assess general metacognitive skills, such as the Metacognitive Awareness Inventory (MAI developed by Schraw & Dennison, 1994) and the Memory and Reasoning Competence Inventory (MARCI, Kleitman & Stankov, 2001). The MAI is a 52-item questionnaire testing people’s awareness of themselves in a non-specific learning context and assessing the awareness of general learning ability and strategies (using such item as “I know how well I did once I finish a test”). Not as extensive as the MAI, the MARCI is used to assess knowledge about one’s own memory and reasoning ability (using items such as “I can remember more material than the average person”). One of the most representative and typical measures in the domain of learning is the think-aloud method, in which records the learner’s ongoing metacognition behavior. Learners are asked to verbalize their thoughts while doing the task. This allows metacognitive behavior to be recorded directly and completely. However, transcribing and judging verbalized thoughts requires adequate time and proper experience of using the method, which limits its popularity (Azevedo et al., 2010; Schellings, 2011). Another proven effective method to get insight into students’ metacognition is to record and assess the actions that learners perform when working on the task. For example, Van Essen and Hamaker (1990) asked students to make a drawing when solving a word problem in order to reveal their strategies of analyzing and exploring a problem. This visualization method has frequently been used in metacognition of problem solving (Edens and Potter 2007; Hegarty and Kozhevnikov 1999; Van Garderen and Montague 2003).

1.2.2 Approaches quantify the accuracy of metacognition

Metacognition accuracy refers to how closely metacognitive judgements track ongoing task performance. A naive approach to metacognition accuracy might insist that we are in complete control of own cognitive behavior. On this view, the self is regarded as “epistemic authority” (Ellis & Kruglanski, 1992). However, Nelson, Kruglanski, and Jost (1998) present a very different view. In their analysis, multiple sources of information and ability influence how people assess knowledge of themselves and others, such as motivation, cultural beliefs and interpreting capacity. Since people may not be natural metacognition experts, measuring metacognitive accuracy has become a major research interest.

There are multiple ways to measure metacognitive accuracy, but the core of these is to compare metacognitive performance with cognitive performance directly. The specific analysis depends on the specific metacognitive measure in each domain. Van Garderen and Montague (2003) used Pearson correlation coefficients to examine relationships between visual-spatial representation (metacognitive skill) and mathematical problem-solving performance. The simple analysis of correlation is one of the most frequently used analysis. Those who used a separate metacognitive scale also use correlation to test whether results on the self-report scale predict the person's performance in another cognitive task. For example, Shah et al., (2015) found strong correlation between the PI20 score (a 20-item prosopagnosia index, a standard self-report instrument assessing the general face perception ability) and performance on face recognition task. However, the drawback of this measure is that the results of a different scale could not reveal how people assess their performance in a particular task. Besides correlation, there is also a common analysis that compares differences in mean performance between the meta level and object level. For example, Kruger and Dunning (1999) used paired sample t-tests to compare participants' estimated percentile ranking and the actual ranking in a logical reasoning task. The results revealed whether participants had overestimated or underestimated their own performance.

In addition to the above typical methods, there are still a variety ways of measuring and analyzing metacognition accuracy, such as the Goodman–Kruskall gamma coefficient, G (Goodman and Kruskal, 1954) which mainly suits the designs using a rating scale in confidence testing and Phi (φ) correlation which mainly suits dichotomous low/high confidence designs (see the comparison of each measurement in Fleming & Lau, 2014). Meta-*d*' (Maniscalco and Lau, 2012) is a more recently developed approach that further quantifies the efficacy with which confidence ratings discriminate between correct and incorrect judgments in a signal detection theory (SDT) framework. However, this is built upon parametric assumptions that some metacognitive measures may not meet. In general, the approach to analyzing metacognitive accuracy depends on the type of metacognition measurements in different domains.

1.3 Metacognitive Illusions

According to above review, people's metacognition accuracy varies. Unfortunately, their unrealized false judgments cause some unexpected outcomes, which has been a focus of research in metacognition for both the insights into oneself (e.g. the Dunning-Kruger effect) and insights into others (e.g. egocentric bias).

1.3.1 Dunning-Kruger Effect

Early metacognition research mainly focused on the field of educational settings (Hacker, Dunlosky, & Graesser, 2009), developmental psychology (Brown, 1977, 1978; Masters, 1981) and clinical psychology (Meichenbaum & Asarnow, 1979; Merluzzi, Rudy, & Glass, 1981). In

1999, the social psychologist Justin Kruger and David Dunning expanded metacognition research into social psychology and observed an intriguing metacognition illusion, now known as the Dunning–Kruger Effect (DKE). Since then, researchers have devoted significant attention to exploring this illusion in various domains. Kruger and Dunning (1999) found that people tend to overestimate their abilities due to the lack of metacognitive skill. They based their conclusions on studies of performance in tests of humor, logical reasoning, and English grammar, combined with participants' self-assessment of their performance in these tests. They found that low performers (those in the lowest quartile) overestimated their percentile ranking (that is, their relative performance compared with other people) as well as their test scores (that is, their absolute performance, the accuracy in the task) by some 40 to 50 points (see Figure 1). These low-performing participants believed that they were outperforming the majority when, in fact, they were the ones being outperformed. Conversely, high performers (those in the highest quartile) were typically more conservative and underestimated their own performance. One important contribution of Kruger and Dunning's analysis is that it explored metacognition performance from the perspective of individual differences. It is a typical finding of individual differences in metacognition showing the difference of metacognitive ability between high and low performers. Some people are more inclined to introspection than others (Stanovich, 2012; Stanovich & West, 1998, 2000). Understanding what makes the difference from those people and those who are unskilled but unaware will contribute to improve the metacognitive ability and further help promote their cognitive outcomes.

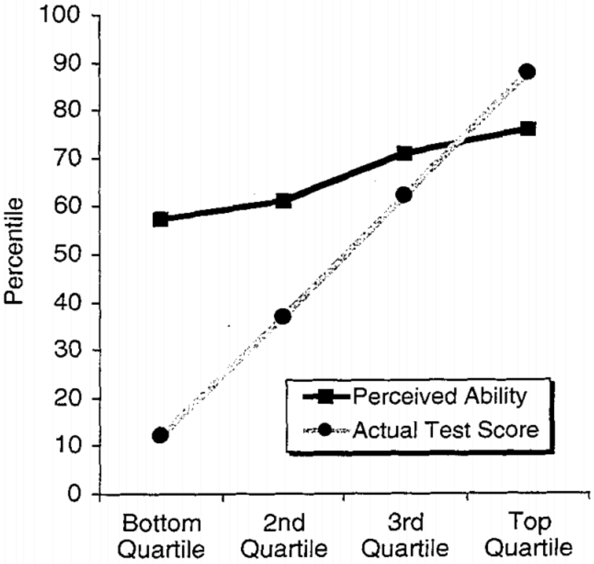


Figure 1.1 The comparison of perceived ability and actual test score across Participant Group (bottom, 2nd, 3rd or top quartile) in humor recognition task (Kruger & Dunning, 1999. Study 1).

Dunning-Kruger's unskilled-and-unaware effect was a landmark finding in how people look at themselves and has been replicated in many research areas. For example, reasoning (Kruger & Dunning, 1999; Pennycook, Ross, Koehler, & Fugelsang, 2017), information literacy skill (Mahmood, 2016) and has even been expanded to political psychology in recent years (Anson, 2018; Motta, Callaghan, & Sylvester, 2018). Furthermore, Dunning, et al., (2003) suggested that it is worthwhile to explore metacognition in other different domains, so that we can determine whether there are any "special" domains in which people correctly intuit their deficits, and how those domains might differ from others.

1.3.2 Egocentric bias

When asking people to see things from another person’s perspective, they tend to estimate it first from their own perspective. This is probably because we have privileged access to information about ourselves (e.g., private thoughts, emotions). We spend much more time paying attention to our own thoughts or feelings than to those of other people. We do not have direct access to the thoughts of other people, so it is difficult to adjust from our original viewpoint to that of others. When people rely too much on their own opinions and fail to properly evaluate the feelings of others, this can lead to an egocentric bias. That is generally understood to mean that “individuals tend to perceive events largely from their own perspective” (Greenwald, 1980). Egocentric bias is usually conceived of as the primary mechanism behind several related cognitive biases, including the tendency to overestimate the degree to which other people notice and care about one’s own appearance and actions (the spotlight effect; Gilovich et al., 2000), the tendency to believe that their opinions and beliefs are more common in the population than they actually are (the false consensus effect; Ross et al., 1977), the tendency to rely too heavily on the first piece of information we are given about a topic (anchoring bias; Furnham & Boo, 2011) and the distinct condition which showed the tendency to believe we are less biased than our peers (Blind Spot Bias; Pronin, 2007). It also plays an important role in fairness perception in which people believe that situations that favor them are fair (Tanaka, 1993). Given the fact that egocentric bias can strongly influence the way we process information and make decisions, it has been widely explored in different areas of psychology (Hinds, 1999; Kelley & Jacoby, 1996; Dimaggio et al., 2008) and has been described as “ubiquitous” (Nickerson, 1998).

Thus, egocentric bias also occurs in the field of metacognition. One concept that is closely related to egocentric bias is theory of mind, the ability to understand other people's mental states (Heyes, 1998). Egocentric bias occurs when people need to use their theory of mind. We tend to look at others in an egocentric way. For example, one study proposing an integrated model of narcissistic personality via clinical practice put this metacognitive deficit at the heart of the model. They found the narcissistic patients had limited ability to understand other people's minds due to their own egocentric bias (Dimaggio, et al, 2002). However, egocentric bias is a cognitive bias which hasn't been put forward explicitly in the field of metacognition. Only the same pattern of finding has been raised in metacognition research. For example, to our knowledge, the only demonstration of egocentric bias in metacognition of face identity is by Ritchie, et al (2015). They asked participants to carry out a paired matching task for facial identity, in which half of the pairs showed familiar faces, and half of the pairs showed unfamiliar faces. For each pair, they also asked participants to indicate how many other people would answer correctly. As expected, people matched images of familiar faces (faces of the staff members from their own university) more accurately than unfamiliar faces (faces of the staff members from another university). Critically, they also predicted that faces they themselves knew would be more accurately matched by others. Nevertheless, the researchers did not use the terms "egocentric bias" or "metacognition" in their face identity study. It is important to understand egocentric bias in metacognition systematically and explicitly in broad fields. Along with its widespread exploration in cognition field, it is plausible to believe that egocentric bias will also be a domain-general phenomenon in metacognition.

1.4 What is face perception

1.4.1 Definition

Face perception refers to the process of understanding and interpreting the face, particularly the human face, especially in relation to the associated information processing in the brain. Face perception is often thought to be ‘special’ in some respects due to its innateness, automaticity, and neural specificity (see, e.g., Farah, Wilson, Drain, & Tanaka, 1998, for discussion). Unlike other abilities that require background knowledge, wisdom or savvy, even newborns can perceive faces and prefer to track face-like stimuli than non-face ones (Johnson, et al., 1991). Just as people often see a face-like patterns in clouds, people detect human faces among other stimuli quite fast with minimum RT around 250-290 ms (Thorpe et al., 1996). Furthermore, personally familiar faces and famous face can be recognized in a few hundred milliseconds (Ramon et al. 2011; Barragan-Jason et al. 2012).

The specificity of facial recognition is usually revealed by comparison to object recognition. For example, it is harder to recognize inverted faces than inverted objects of other categories (Yin, 1969). Unlike in object recognition, people can show a special holistic advantage in identifying individual face parts presented in the entire face than they were at identifying the same part presented in isolation (Tanaka & Farah, 1993). Holistic processing has been proposed as a general mechanism that supports face detection (Taubert et al. 2011). Many behavior studies suggested that there are unique visual mechanisms for face perception (see McKone & Robbins, 2011, for a review based on the evidence from fMRI, ERPs, TMS, and neuropsychology studies).

For example, neuropsychologists have demonstrated the fusiform face area (FFA) is a part of the human visual system that is specialized for facial recognition (Kanwisher et al. 1997), the Occipital Face Area (OFA) is responsible for identifying parts of faces, such as eyes, nose, and mouth (Pitcher et al. 2011) and the posterior superior temporal sulcus (pSTS) is a face-selected region processing the dynamic aspects of faces, such as facial expression and eye gaze (Allison et al., 2000).

1.4.2 Theoretical and practical importance

The study of face perception used to be a minority interest as a sub-branch of visual perception during the early years of cognitive psychology (Calder et al., 2011). Over the past few decades, face perception has become a prolific area of visual research, leading to substantial advances in our understanding at the behavioral, conceptual, neuroscience, developmental and computational levels (Oruc et al., 2019). The fundamental work from Bruce and Young (1986) developed a theoretical model for face recognition (see Figure 1.2). This framework also helped to understand the recognition of word and other objects by parallel comparison with facial recognition. Conceptually, the study of face perception has also contributed to the understanding of the voice, which can be considered as a kind of 'auditory face' (Young et al., 2020). Indeed, many studies have examined face-voice integration in person perception (e.g., Campanella & Belin, 2007; Freeman & Ambady, 2011). Turning to the theoretical contribution at the neuroscience level, besides the finding of face-specific area in the brain, face-selective cells have been found in the temporal cortex of monkeys (Gross et al. 1972). Based on this, Desimone (1991) presented some possible reasons for the evolution of face neurons and suggested some analogies with the development of language in humans. Face perception from a developmental perspective has received a great deal of attention (Dahl et al. 2013; Aylward et al., 2005; de Heering et al., 2012; Simion & Giorgio, 2015), including face perception in children with Autism and Asperger's Syndrome (Davies et al. 1994) or those with developmental prosopagnosia (Corrow et al., 2019; Dalrymple et al., 2014). With rapid advances in computer vision (e.g., Moon and Phillips 2001), research into automatic face perception has also become a burgeoning field (Blauch et al., 2020; Martinez, 2017; O'Toole, 2011; Bartlett & Whitehill, 2010; Bonnen et al., 2013).

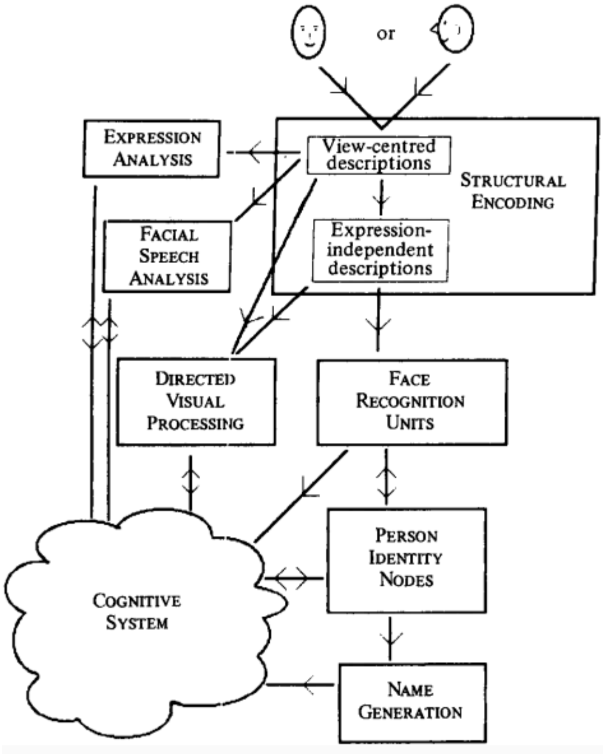


Figure 1.2 A functional model for face recognition (Bruce & Young, 1986).

Given that the face is an important site for the identification of people and conveys significant social information, the practical importance of face perception extends to social interaction, clinical treatments and security and forensic procedures. Rapidly recognizing familiar people, learning new faces and observing others' emotions and social signals from their faces is a core human ability, which is also crucial for interpersonal communication. The study of face perception has been a source of inspiration for social neuroscience (Haxby & Gobbini, 2007) and social psychology (Zebrowitz & Montepare, 2008). Deficits in face perception make it hard to extract cues for appropriate social behaviors from others and are often associated with disorders of social cognition (Gage & Baars, 2018). Unfortunately, some people fail to develop normal face perception abilities (developmental prosopagnosia; Susilo & Duchaine, 2013) and deficits can also follow brain injury (acquired prosopagnosia; Damasio et al., 1982). Over the last 50 years, face perception researchers have demonstrated some effective treatment approaches for both kinds of prosopagnosia (see DeGutis et al., 2014, for the review). The face also serves as one of the major biometric source. The most important advantage of face over biometric modalities such as fingerprint and iris is that it can be captured at a distance (Wheeler et al. 2011) and in a covert manner without awareness (Haan et al., 1987), which make it key to many security and forensic procedures, such as surveillance applications (Wheeler et al. 2010). The extensive research on unfamiliar face matching has contributed much to ID checking in real-life scenarios. For example, Wirth and Carbon (2017) explored how professional experience and time pressure impact on passport-matching performance according to a series of passport face matching studies. Increasing face memory research has showed great implications for eyewitness testimony. For example, Grabman et al. (2019) found that recognition ability, decision-time, and justifications all cause high-confidence errors based on a lineup eyewitness memory task. So they suggested the justice system should consider both individual differences and confidence in determining whether an eyewitness's decision is likely to be accurate. Findings of psychology of face perception have also been used by artificial intelligence to inform software design which provides for brain-machine interface for facial recognition (Njemanze, 2004). Also, a variety of tests were created by face researchers to help with the selection of super-recognizers (SRs)— professionals trained in unfamiliar face recognition for security-critical roles. For example, the UNSW Face Test is a screening tool for super-recognizer designed by the researchers from UNSW (Dunn et al., 2020).

1.5 Face perception measurements

Faces convey rich information about the people around us. The details of what kind of information we derive from face will be introduced later in this chapter. Here I focus on ways to measure cognitive tasks involving faces – recognizing identity, recognizing emotion, interpreting gaze, remembering new faces, perceiving social characteristics and so on.

1.5.1 Face recognition

One of the most common measures of face identification is face matching. Bruce et al, (1999) created a line-up matching task that has been widely used in face matching studies. Participants were shown a target face image captured from a video, together with an array of 10 high-quality faces with similar physical appearance to the target. Half of the trials were target-absent arrays and half of the trials were target-present arrays (see Figure 1.3 for the examples of each condition). Participants were asked to decide whether or not the target was present, and if so, which face it was. This kind of 1-in-10 matching is commonly encountered in the eyewitness identification.

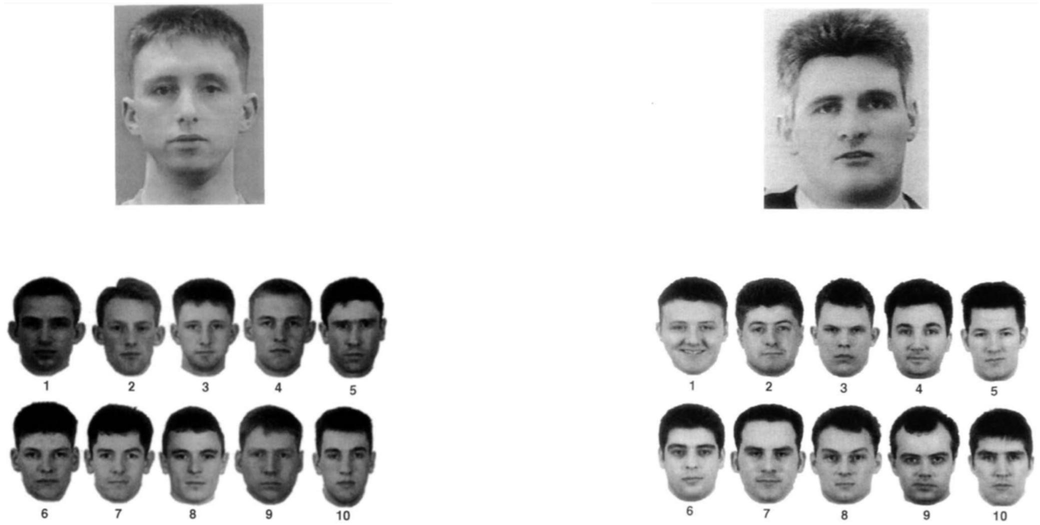


Figure 1.3 Two example trials from the line-up matching experiment of Bruce et al., 1999. Left panel shows a target-present condition with the target at the array number 3. Right panel shows a target-absent array.

Getting 10 faces who look like one target sometimes seems hard. Also, all the faces were taken from one video filmed on the same day. However, faces are usually captured by different cameras encountered no matter in the security settings and ID checking. Burton, White and McNeill (2010) developed a new unfamiliar face-matching test, the Glasgow Face Matching Test (GFMT). It simplified the procedure to a paired matching task, in which viewers were presented with pairs of unfamiliar faces in similar frontal pose but taken with different cameras (see Figure 1.4 for examples). The task is simply to decide whether the two images showed a same person (matching pair) or two different people (mismatching pair).

Figure 1.4 Two example trials from the Glasgow Face Matching Test. Left panel shows a mismatching pair. Right panel shows a matching pair.

Considering the wide range of variability in photos of the same person, Jenkins et al. (2011) described a face sorting measure in which participants were shown 40 different face photos (see Figure 1.5) and asked to group the photos of the same person together. Participants were free to create as many or as few groups as they wished. The correct solution is two groups (two identities). Participants who are familiar with the two identities performed perfectly but those who are not often create 7 or 8 different groups.

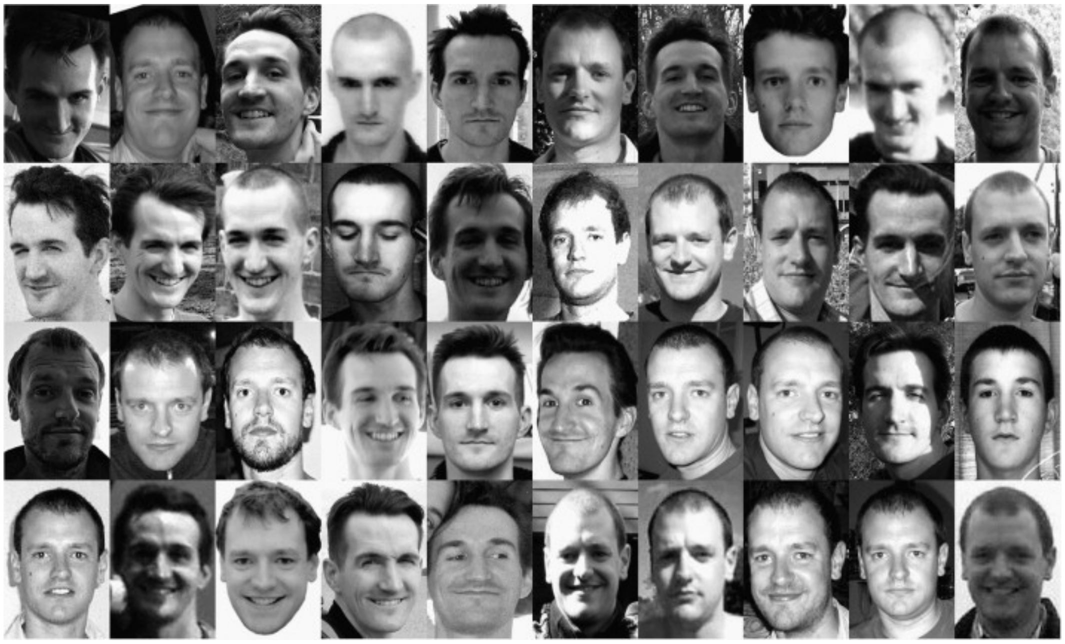


Figure 1.5 The 40 face photos (20 for each identity) from Jenkins et al.'s (2011). Solution: (Row 1) ABAAABABAB, (Row 2) AAAAABBBAB, (Row 3) BBBAAABBAA, (Row 4) BABAABBBBB.

Ekman's 60 Faces Test has proven to be a sensitive test for recognition of emotional expressions on faces (Young et al., 2002). Participants were shown 60 face images, consisting of 10 models (4 male, 6 female), each expressing six different facial expressions (Fear, Disgust, Anger, Happiness, Sadness or Surprise, see Figure 1.6). After looking at each image, participants were asked to choose which emotion is being displayed. The total score ranges from 0–60, with higher scores indicating higher ability to recognize the emotions in human faces.

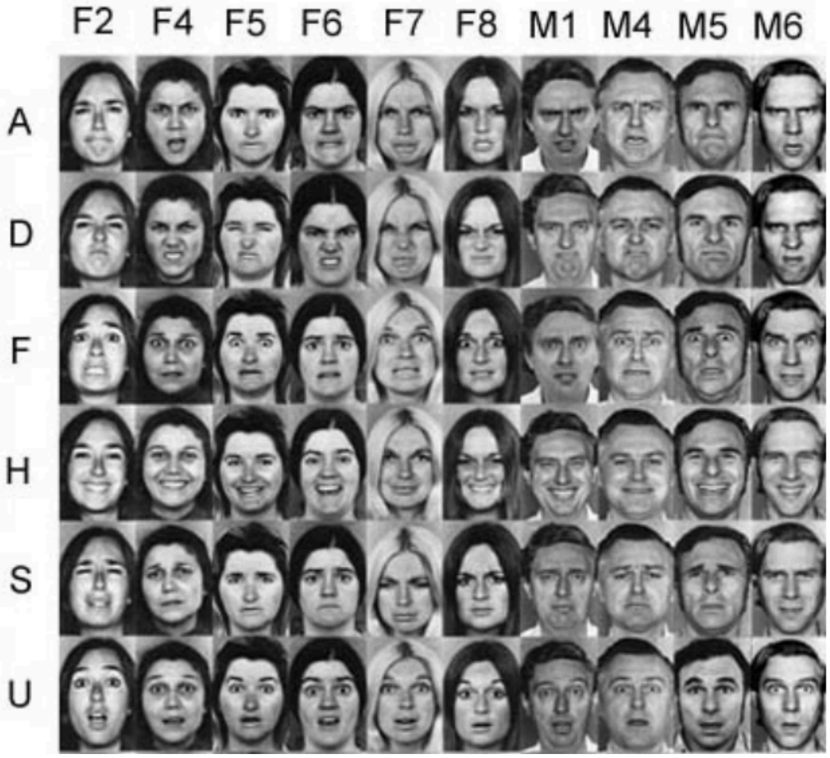


Figure 1.6 The 60 face photos (6 for each model) from Ekman and Friesen (1976) Pictures of Facial Affect used in FEEST. There are emotional expressions of anger (A), disgust (D), fear (F), happiness (H), sadness (S) and surprise (U) pose for 6 female and 4 male models. The labels used to identify each of the models locate them in the Ekman and Friesen series (F2 = second female model in the series, M1 = first male model, etc.).

Gaze direction is usually regarded to modulate or influence face identity, emotion recognition or face memory. Thus researchers often used the eye-tracking system (e.g., Ho, Foulsham & Kingstone, 2015; Laidlaw & Kingstone, 2017) created their own gaze images stimuli according to the specific research aims and apply the factor of gaze with the existed paradigm in other field (e.g., Smith et al., 2006; Calder et al., 2002; Ueda & Koyama, 2011). One typical measure is from Jenkins et al. (2006) who used an adaptation paradigm to investigate the functional organization of gaze perception. They created a set of 60 face images (12 models posing five different angles of gaze) with staring or deviated eyes (see Figure 1.7 for the examples). In their gaze acuity test, participants were shown a series of face images. For each task, they were asked to make the eye gaze judgement (left, straight or right).

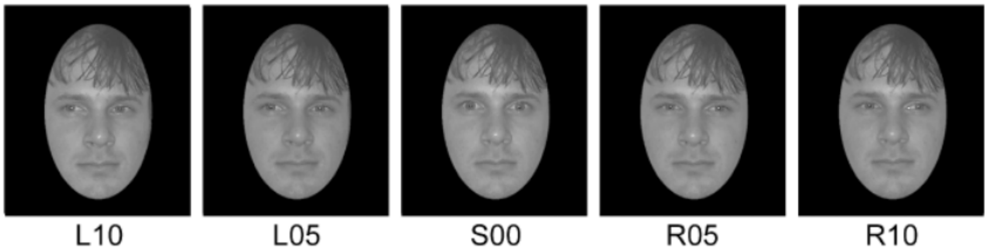


Figure 1.7 Example stimuli of the five gaze angles used for the gaze-acuity test from Jenkins et al., (2006). L10 = 10° left; L05 = 5° left; S00 = straight ahead; R05 = 5° right; R10 = 10° right.

1.5.2 Face memory

The main paradigm of the tradition face memory task is consisted with two phases: the learning phase and the testing phase. Participants were asked to learn some faces in the learning phase and then test whether they can recognize the old faces they learned in the learning phase from the new faces. This old/new face memory paradigm has been widely applied (e.g. Zaki &

Johnson, 2013; Tüttenberg & Wiese, 2019; Hourihan et al. 2012; Hills, 2012) and some new tools building on this paradigm have also been developed. One of the most widely used tests is the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). In the learning phase of this test, participants are asked to learn six target faces, each in three views. The test phase contains three stages of 1-in-3 forced choice: three trials test the recognition of the identical image to the learning phase, five trials test the recognition of the same faces in different images (different viewpoint and/or lighting) and four trials test the recognition of the same faces in different images covered with heavy visual noise (see Figure 1.8).

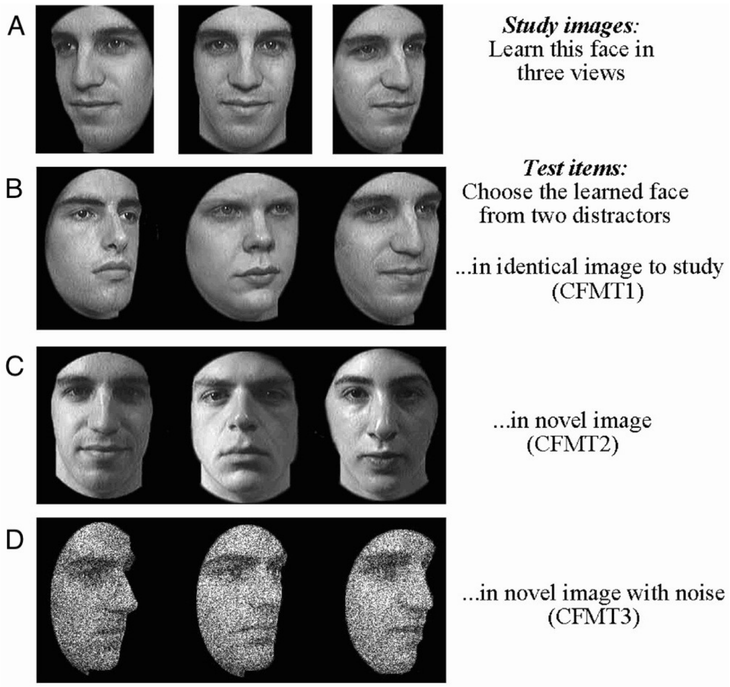


Figure 1.8 Examples of study images similar to test stimuli from the Cambridge Face Memory Test. None of these items was used in the test. Panel A shows the three views of a target face in the learning phase. Panel B displays a test trial from the same image stage (the rightmost image is the target). Panel C shows a test trial from the novel image stage (the leftmost image is the target). Panel D displays a test trial from the novel images with noise stage (the rightmost image is the target).

1.5.3 First impressions

The accuracy of the trait inferences from facial appearance is usually tested via correlations with self-reports. For example, Penton-Voak et al. (2006) collected neutral pose photographs from 294 volunteers who also completed a self-report personality questionnaire (Botwin, Buss, & Shackelford, 1997). The 294 photographs were then rated on the Big Five personality dimensions (Agreeableness, Conscientiousness, Extraversion, Emotional Stability, Openness) by 10 raters on a 1-7 scale. However, people always associated lots of traits not limited to the several traits the researcher listed. Sutherland et al. (2018) used an open question paradigm. Participants were shown a series of faces one at a time and were asked to type in anything that came to mind upon viewing the face. Afterwards, two colleagues were asked to categorize all the content participants gave to each face. The main categories were then used as the dimensions in a 1-7 personality rating scale.

1.6 What can we derive from face?

From looking at someone's face we can recognize whether we know this person, we can infer their mood state, the direction they are looking at, their racial background, whether they look trustworthy or untrustworthy, attractive or unattractive, and so on. Besides, there are still lots of information face conveys, such as messages from facial movements (e.g., lipreading). In this section, I review the major categrories of information that we viewers derive from faces.

1.6.1 Face identity

In daily life, we see many unfamiliar faces such as the strangers on the way home. We also recognize people we are already familiar with, such as friends , family, colleagues, as well as media representations of celebrities. It is now well established that familiar and unfamiliar face perception are largely different (see Hancock et al., 2000 for the review). Bruce and Young (1986) suggested that unfamiliar faces were principally processed by the use of pictorial codes while structural codes played an important on familiar faces processing. This has been revisited by Hanley, Pearson, and Young (1990), who suggested that familiar and unfamiliar faces were both processed based on pictorial memory but with separate systems. Based on the difference in processing, people performed differently in familiar and unfamiliar face identification.

In recent years, it has become clear that our ability to identify unfamiliar faces is surprisingly poor on average. For example, participants were asked to pick out one out of 10 full-face photographs that they thought matched the single unfamiliar target face which had been taken from a video filmed on the same day (see Figure 1.3). Even though the photographs were all in high quality, their performance was still highly error prone, with a range from 30% to 39% error in each target pose condition (Bruce, et al., 1999). This was consistent with the results from the Glasgow Face Matching Test (GFMT), with around 20% error rate on average (Burton, White, & McNeill, 2010). The GFMT also revealed very large individual differences in performance, with 100% accuracy for high performers but only 62% for the low performers. This finding has been replicated in many other studies (e.g. White, et al., 2014). Together, these studies reveal a broad range of face recognition ability. At one extreme, developmental prosopagnosics show profound impairments in face identification specifically (Duchaine & Nakayama, 2006). At the other extreme, 'super-recognisers' show exceptional ability that has clear operational value (Robertson, et al., 2016). The Dunning-Kruger effect was a typical finding of individual differences in metacognition showing the difference of metacognitive ability between high and low performers (Kruger & Dunning, 1999). However, very little is known about people's insight into their own face perception abilities, or those of other people. If recognizers are "unskilled and unaware of it", this could explain society's enduring confidence in facial image comparison as method of identifying unfamiliar people, despite evidence of its unreliability.

For familiar faces, the situation is very different. Familiar face recognition refers to the recognition of famous face such as celebrities or personally familiar people such as family and friends. In contrast to strangers' faces, viewers are extremely good at identifying faces that they already know. For example, Burton et al. (1999) showed participants the video clips of 20 members of the lecturing staff in a particular department. Half of the participants were students in the same department as the staff (familiar group) and the other half were students from a different department (unfamiliar group). They then presented 20 high-quality images, telling participants that half of these faces had been shown earlier, and asked participants to indicate their confidence that this person was in the video. The results showed that the familiar group performed much more accurately than the unfamiliar group. Even when the researchers used poor quality video footage, the familiar group kept performing well. It seems that familiar face recognition is an easier task than the unfamiliar face recognition. One open question is whether people's metacognition performance is also better in the familiar face task, and whether the

Dunning-Kruger Effect will still exist even there is no big individual difference in their cognitive performance.

1.6.2 Social signals

Besides face identity, there are also other important social aspects of variant information that people can extract from the face, such as emotional expression and gaze direction. There are thought to be six basic or universal emotions: anger, disgust, fear, happiness, sadness and surprise (Young, et al., 2002). Numerous studies have shown that happiness is better recognized, compared with other negative emotions, especially fear (Calvo & Lundqvist, 2008; Ekman, Friesen, & Ellsworth, 1972). Interpreting people's emotional expression correctly is vital when deciding how to respond to them. Correct assessment of others' expressions can help to reduce the waste of time of guessing other people's feelings and avoid embarrassment in social interaction. Based on different study designs and measurements of metacognitive ability, researchers have drawn different conclusions about individual differences in metacognition for emotional expression. For example, Ickes et al (1990) asked participants to revisit video footage of their interactions with another participant, then recognize the emotion exchanges and later fill in a self-report questionnaire of empathy. They found that people who reported high empathy did not show higher accuracy at emotion recognition. Conversely, Kelly & Metcalfe (2011) suggested that people who were good at recognizing emotional expression also had good metacognition knowledge by measuring the gamma correlation between metacognition and performance in the Mind in eye task (Baron-Cohen et al., 2001). Gaze direction derived from faces can also serve as a social signal with multiple meanings. Another person's eye gaze can direct our spatial attention (Friesen & Kingstone, 1998) or signal the referent of a remark (Hole & Bourne, 2010). Baron-Cohen (1997a, 1997b) noted that children's ability to detect gaze direction is highly important in the development of Theory of Mind (i.e., one aspect of metacognition). Nevertheless, much less has been done in exploring metacognition in gaze direction recognition.

Furthermore, there are some potential interactions among face identity, emotional expression and gaze direction. Although the initial model of face processing suggested that the processing of face identity and emotional expression are independent (Bruce & Young, 1986), the interdependent view has gathered some support in recent years. For example, Kaufmann and Schweinberger (2004) found that emotional expression influenced identification of familiar faces, with faster recognition for happy faces. Gaze direction has also been found to interact with emotional expression. For example, combining the fearful faces and averted gazes resulted in an adaptive fear response (gaze shifting) to danger (Hadjikhani et al., 2008). Since face identity, emotional expression and gaze direction may interact with each other in the cognition level, their relationship in the metacognition level should also be interesting to focus on.

1.6.3 Social categorization

Faces also convey information about social categories. At the basic level, they indicate species, which is important not only for humans but also in a wide range of animals (Leopold & Rhodes 2010). Viewers also read cues to gender, age and race from faces. A number of studies have found that people perform better at recognizing faces that are similar to their own (Minear & Park, 2004). The effect has been found for gender and age. For own-gender bias, researchers have found that females typically outperform males in face recognition tasks (Herlitz & Yonker, 2002). Other studies have reported a significant interaction between gender of subject and gender of items, with both male and females identifying their own gender best (Mason, 1986). Own-age bias refers to the finding that viewers better recognize faces of their own age group than faces of another age group (e.g., Rhodes & Anastasi, 2012). Compared with gender and age, race effects in face memory are the best-known phenomenon. The Other-Race Effect (ORE; also known as the Cross-Race Effect, Own-Race Bias or Other-Race Deficit) is a robust phenomenon in face recognition (see Meissner & Brigham, 2001 for a review).

To date, the explanations for the ORE have delivered somewhat contentious findings which can be summarized into two representative models (see Shriver et al., 2008 for a review). An early hypothesis is that people have greater expertise with own-race faces than other-race faces due to the lack of contact with other-race people (see MacLin & Malpass, 2001). This is known as the Perceptual Expertise Model. One implication of this model is that increase contact with outgroups can improve other-race face recognition. However, how much contact can be effective is hard to measure based on different situations and experiment settings. This model can be understood as supporting the view that familiarity plays an important role in face memory. An alternative view, the Social-Cognitive Model, claims that people encode outgroup faces categorically (e.g., race, gender, age) but encode ingroup faces individually (i.e., face features). That is, even if you are familiar with other-race people, you still encode their faces according to their race rather than their face features, leading to poorer face memory. This model emphasizes categorization, rather than individual familiarity as a determinant of face memory.

Even though familiarity is related to both of these theories, the potential role of familiarity has received little attention as a contributing factor to face memory. Much less has been done in comparing the effect of familiarity and race on face recognition. Because previous face recognition research has mainly focused on race rather than familiarity. For example, Zhou & Mondloch (2016) asked participants to sort photographs of two identities into as many piles as they thought which stand for the same person. The unfamiliar face sorting results showed significant ORE, with more piles sorted for other-race faces than own-race faces. While for familiar faces, participants made correct two piles for both own- and other-race faces. Neuropsychologist demonstrated the same conclusion that familiarity can eliminate the ORE. Phelps et al. (2000) experiment 1 and Hart et al., (2000) found that amygdala showed greater activation to black than white faces when white participants performed an unfamiliar face matching task. To test the familiarity effect, Phelps et al. (2000) showed white participants familiar faces from both races in experiment 2 and found no ORE in amygdala activation. Barzut et al. (2013) appear to be the only researchers to have examined the two factors in the old/new face memory paradigm. They found the ORE did decrease after the introduction of familiar face, but it still existed. So they claimed that familiarity has been put too much weight on. However, this conclusion was obtained only from one race of participant. Also, all prior studies explored whether familiarity would reduce ORE but none has considered familiarity and race as two independent factors. Since people are better at recognizing own-race faces than other-race faces and our memory of familiar faces is also highly accurate (Ge et al., 2003) compared with unfamiliar faces (Nakashima et al., 2012), it may be informative to establish which factor is more psychologically important at the level of memory performance.

Also of interest is whether people realized their bias, that is how people think the two factors would influence the face memory performance of themselves and that of other people including the ingroup and outgroup peers. The metacognition aspect of face memory is of highly practical value in the context of the criminal justice system, especially the testimony of the eyewitness in verdict decisions (Benton et al., 2006). Eyewitnesses have some insight into their own performance in identifying own-race or other-race suspects, and their testimony is also valued by jurors, whose race might either be the same as or different from that of the eyewitness (Neil v. Biggers, 1972). The suspect could also be a familiar face to the eyewitness, such as a neighbor, friend, or colleague. Applied face recognition underscores the importance of understanding metacognition for face memory, and for examining race (own- or other-race faces) and familiarity (familiar or unfamiliar faces) conditions together. Previous research into metacognition for the ORE (e.g. Hourihan et al., 2012; Rhodes et al., 2013) found that white participants were more confident in their performance when identifying an own-race suspect than an other-race suspect in a line-up (Smith et al. 2004). However, it is not known whether people will show this bias in familiar face memory or how they expect face race and face familiarity to affect other people's performance.

1.6.4 Subjective inference

In addition to objective information, people also make subjective inferences from faces. For example, choosing people who look friendly when asking for directions on the street. Despite the fact that personality inferences from facial appearance have long been studied (Hollingworth, 1922), there is little consensus as to whether people make accurate trait inferences from faces. Some studies have reported positive correlations between trait inferences and self-reports of approachability, warmth, power and extraversion (Berry, 1991; Berry & Brownlow, 1989; Penton-Voak et al., 2006). Other studies have failed to find such relationships for agreeableness, conscientiousness, and suggestibility (Pound et al. 2007; Bachmann & Nurmoja 2006). According to Todorov et al. (2011), one important reason for these inconsistent results concerns the difficulty of ascertaining the representativeness of the face stimuli. One explanation for this inconsistency is that the accuracy depends on different trait dimensions. Oosterhof and Todorov (2008) introduced a two-dimensional model of social inferences based on trustworthiness and dominance (based on computer generated images). Recently, Sutherland et al. (2013) extended the model to three dimensions: trustworthiness, dominance and attractiveness (based on real face photographs). In addition to the first impression of personality inferred from the unfamiliar faces, there are many other thoughts that come to mind, such as semantic associations and episodic memories. However, much less has been done in understanding that kind of information.

1.7 Overview of current work

Face perception and metacognition are both key components of humans' cognitive functioning in theory. They also play an important role in social interaction, but previous studies have rarely made the connection between them. This potentially indicates an interesting case to test whether the same metacognitive principles apply to face perception as to other aspects of cognition that have been studied. In addition, face perception is such a special domain that allows me to respond to the recent calls for exploring the Dunning-Kruger effect in a special domain by Dunning, et al., (2003). By modifying established measures of face perception and combining them with metacognition measures, the following experimental chapters in this thesis deal with different aspects of face perception from the point of view of metacognition.

Chapter 2 begins with Experiment 1, which tests whether Dunning-Kruger effects and egocentric bias apply to face identification. Experiment 2 extends this approach to other face perception tasks. The same format of same/different matching tasks was used to test familiar and unfamiliar face identity, emotional expression, and gaze perception. Cognitive measures were combined with metacognitive estimates of test score and percentile ranking measures.

Chapter 3 examines familiarity effects and other-race effects in face memory at both cognitive and metacognitive levels. As noted in the historical review of Metcalfe & Dunlosky (2009), initial work on metacognition is rooted in the memory study. I combined the tradition old/new face memory paradigm with the prospective, concurrent and retrospective metacognition judgements of oneself and others. Experiment 3 used the same images in the learning and test phase. Experiment 4 used different images at learning and test.

Chapter 4 examined egocentric bias in face-evoked thoughts using a new face association task modified from the word association task and the trait association task of unfamiliar face from Sutherland et al. (2018). This approach is more widely applicable to both familiar and unfamiliar faces. Experiment 5 examined anything people think about when looking at a face. Experiment 6 examined differences between familiar and unfamiliar faces. Experiment 7 focused on who comes to mind when looking at a familiar or unfamiliar face.

Chapter 2

Dunning–Kruger Effects in Face Perception

Reference:

Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. *Cognition*, *203*, 104345. doi:10.1016/j.cognition.2020.104345

Abstract

The Dunning–Kruger Effect refers to a common failure of metacognitive insight in which people who are incompetent in a given domain are unaware of their incompetence. This effect has been found in a wide range of tasks, raising the question of whether there is any ‘special’ domain in which it is not found. One plausible candidate is face perception, which has sometimes been thought to be ‘special’. To test this possibility, we assessed participants’ insight into their own face perception abilities (self-estimates) and those of other people (peer estimates). We found classic Dunning–Kruger Effects in matching tasks for unfamiliar identity, familiar identity, gaze direction, and emotional expression. Low performers overestimated themselves, and high performers underestimated themselves. Interestingly, participants’ self-estimates were more stable across tasks than their actual performance. In addition, peer estimates revealed a consistent egocentric bias. High performers attributed higher accuracy to other people than did low performers. We conclude that metacognitive insight into face perception abilities is limited and subject to systematic biases. Our findings urge caution when interpreting self-report measures of face perception ability. They also reveal a fundamental source of uncertainty in social interactions.

2.1 Introduction

Negotiating everyday life requires that our plans are commensurate with our abilities. This basic requirement underscores the importance of metacognition—insight into one's own thinking and the thinking of others (Fleming, Dolan, & Frith, 2012; Jost, Kruglanski, & Nelson, 1998; Tullis & Fraundorf, 2017). In fact, metacognitive insight is not only inaccurate, it is also subject to systematic biases. One influential example of such a bias is the Dunning–Kruger Effect, encapsulated in the title of its debut paper, "unskilled and unaware of it" (Kruger & Dunning, 1999). The headline result is that poor performers in a given task drastically overestimate their ability, believing that they are outperforming the majority when, in fact, they are the ones being outperformed (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Kruger & Dunning's (1999) explanation of this effect is elegant, and points to a cruel trap in human psychology: The skills that grant competence in a particular domain are the very skills needed to evaluate competence in that domain. People who lack the former lack the latter. A secondary result concerns the top of the ability range. High performers often *underestimate* their standing, but for an entirely different reason. These individuals recognise that they perform well, they just assume that other people perform well too.

Part of the appeal of the Dunning–Kruger Effect is its broad generality. The same basic pattern emerges in reasoning (Kruger & Dunning, 1999; Pennycook, Ross, Koehler, & Fugelsang, 2017), humour (Kruger & Dunning, 1999), political knowledge (Anson, 2018; Motta, Callaghan, & Sylvester, 2018), and many other domains. Indeed, the apparent ubiquity of Dunning–Kruger

Effects has prompted some to wonder if there is any ‘special’ domain in which the standard pattern is not found (Dunning, Johnson, Ehrlinger, & Kruger, 2003).

One plausible candidate for such a ‘special’ domain is face perception. Evidence that faces may be cognitively special comes from at least four sources (McKone & Robbins, 2011). First, developmental studies have suggested that newborns demonstrate some innate knowledge of facial structure (Goren, Sarty, & Wu, 1975; Johnson, 2005). Second, face perception seems to be disproportionately affected by image manipulations such as inversion (Yin, 1969; Rossion, 2008) and contrast reversal (Kemp, Pike, White, & Musselman, 1996; Farroni et al., 2005). Third, it has been proposed that face perception may be subserved by face-specific neural circuitry (Riddoch et al., 2008; Kanwisher & Yovel, 2006). More recently, genetic studies have shown that human face recognition ability is specific and heritable (Wilmer et al., 2010; Wilmer 2017). This converging evidence from highly diverse studies has led some researchers to propose that face perception may involve specialised or face-specific cognitive processes.

Despite the theoretical and applied interest in face processing, no previous studies have tested for Dunning–Kruger Effects in this domain. A few studies have found that individuals in the general population show minimal to moderate insight into their own face recognition abilities (e.g. Bindemann, Attard, & Johnston, 2014; Palermo et al., 2017; Bobak, Mileva, & Hancock, 2019), echoing findings for other types of memory (Beaudoin & Desrichard, 2011; but see Livingston & Shah, 2018; Arizpe et al., 2019 for more positive views). However, none of these studies was concerned with metacognition in the ‘expansive’ sense that includes insight into other people’s abilities (Jost, Kruglanski, & Nelson, 1998). Their main interest was whether a person’s self-report (e.g. agreement with questionnaire items such as, “My face recognition ability is worse than most people”; Shah et al., 2015) could predict the same person’s performance on standard face recognition tests. They did not compare estimated performance and actual performance for the same task.

A few face perception studies have examined other aspects of metacognition. Sauerland et al (2016) adapted the choice blindness paradigm (Johansson, Hall, Sikström, & Olsson, 2005) to investigate insight into identification judgements. Participants were asked to sort photographs of unfamiliar faces by identity (Jenkins, White, Van Montfort, & Burton, 2011). They were then confronted with one of their identity decisions and asked to justify it. On critical trials, the photographs were secretly switched, so that the decisions participants were asked to justify were opposite to the decisions that they actually made. Very few of these manipulations were detected. Indeed, participants readily reported their reasoning behind identity decisions that they had not reached.

Such findings suggest that insight into one’s own face recognition performance is somewhat limited. Fewer studies have examined insight into other people’s face recognition performance. Ritchie et al. (2015) presented pairs of faces in a matching task for identity. As expected, participants performed better with familiar faces than with unfamiliar faces (Clutterbuck & Johnston, 2004; Noyes & Jenkins, 2017, 2019). However, participants also predicted that the faces they themselves knew would be easier for other people to match—even people who did not know those faces. These findings demonstrate an egocentric bias in identification performance (Greenwald, 1980), in that viewers estimated the cognition of others from their own perspective (DiMaggio et al., 2008; Hinds, 1999; Kelley & Jacoby, 1996). However, it remains unclear whether high-performing participants produced higher estimates than low-performing participants.

Given that face perception is often presented as a special case for cognition, we tested whether it is a special case for metacognition. Specifically, we asked whether standard Dunning–Kruger Effects and egocentric bias emerge in face perception tasks. We begin in Experiment 1 with identification tasks for familiar and unfamiliar faces. In Experiment 2, we expand our analysis to include other aspects of face perception, namely gaze direction, and emotional expression.

2.2 Experiment 1 Identity matching for familiar and unfamiliar faces

Our first experiment had two main aims. First, we sought to establish whether face perception follows the same metacognitive principles as other aspects of cognition. Specifically, we asked whether Dunning–Kruger Effects and egocentric bias are observed in face identification tasks. Second, we sought to compare these metacognitive patterns for familiar and unfamiliar faces. To address these questions, we adapted a standard perceptual matching task for facial identity (Burton, White, McNeill, 2010). In the standard task, participants are presented with pairs of face photos. For each pair, the task is to decide whether the two photos show the same person (50% of trials) or different people (50% of trials). Accuracy on this task is typically at ceiling for familiar faces (e.g. Clutterbuck & Johnston, 2004; Noyes & Jenkins, 2017, 2019), but is generally much lower for unfamiliar faces (e.g. Clutterbuck & Johnston, 2004; Burton, White, McNeill, 2010; Noyes & Jenkins, 2017, 2019).

This task has several characteristics that make it well suited to comparison of cognition and metacognition. First, each trial has a correct answer, so accuracy can be scored objectively. Second, the same/different response options mean that ceiling performance and chance performance are well defined (100% accuracy and 50% accuracy respectively). Third, there are large individual differences in performance (White, Kemp, Jenkins, Matheson, & Burton, 2014), such that high- and low-scoring respondents tend to be clearly separated. Recording actual scores allows us to assign participants to performance quartiles, as per Dunning & Kruger (1999). Recording participants' estimated scores allows us to test (i) whether 'incompetent' participants (lowest performance quartile) show the classic 'unskilled and unaware' pattern, and (ii) whether 'competent' participants (highest performance quartile) underestimated their performance.

Previous studies of metacognition have often relied on retrospective estimates of performance, collected after the whole task (e.g. Dunning & Kruger, 1999; Tenenberg & Murphy, 2005; Simons, 2013; Feld, Sauermann & de Grip, 2017; see Sarac & Karakelle, 2012; Gignac & Zajenkowski, 2020, for useful discussions of this issue). That approach has several drawbacks. One is that it imposes substantial demands on retrospective memory. Cognitive tasks often involve dozens of trials or items, and these will typically vary in subjective difficulty. The challenge is not only to recall the landscape of that experience, but also to encapsulate it in a single score. To complicate matters, the overall impression may be skewed by primacy and recency effects (Haugtvedt & Wegener, 1994). To avoid these issues, we captured (i) actual performance, (ii) self-estimates, and (iii) peer estimates on each trial. We also captured participants' self-estimates of their own percentile ranking at the end of each task, for 'backward compatibility' with previous studies.

In light of Dunning & Kruger's (1999) findings, we predicted that low performers would overestimate their performance and that high performers would underestimate their performance. Since virtually everyone is a high performer for familiar face identification (Burton, Wilson, Cowan, & Bruce, 1999; Jenkins & Kerr, 2013), we expected this interaction to be compressed (near ceiling) for familiar faces. In light of the egocentric bias (Ross, Greene, & House, 1977), we expected low performers to make low peer estimates, and high performers to make high peer estimates. It follows that peer estimates should be lower for unfamiliar faces than for familiar faces.

2.2.1 Method

Participants

Sixty-four UK students (44 female, 20 male; mean age 20 years; age range 18–26 years) from the University of York took part in exchange for a small payment or course credit. The experiments in this study were approved by the Ethics Committee at the University of York. All participants provided written informed consent.

Stimuli and apparatus

Ambient images of 20 familiar faces (e.g. UK and US celebrities; 10 female, 10 male) and 20 unfamiliar faces (e.g. celebrities from other countries; 10 female, 10 male) were downloaded from online sources. Each image was cropped and resized to 570 pixels high × 380 pixels wide for onscreen presentation. For Different Person trials, we paired faces that resembled each other and matched the same basic verbal description (e.g. young woman with red hair). To avoid image repetition, we collected four photos of each face—two for use in Same Person trials, and two for use in Different Person trials. Each face appeared in Same and Different trials equally often, and each participant saw each image exactly once. To ensure that all participants received identical tasks, all participants received identical image pairings. Experiments were run using a 21.5-inch iMac with i5 processor. Stimulus presentation and data collection were controlled by PsychoPy2 v1.82.00 (Peirce, 2007, 2008).

Design

All participants completed both the *Familiar* and the *Unfamiliar* face matching task in separate blocks. Block order was counterbalanced so that half of the participants encountered the *Familiar* condition first, and half of them encountered the *Unfamiliar* condition first. Within each block, the 40 trials (20 *Same* person, 20 *Different* person) were presented in a random order. All participants contributed the same measures in both tasks—actual performance, self-estimates, and peer estimates.

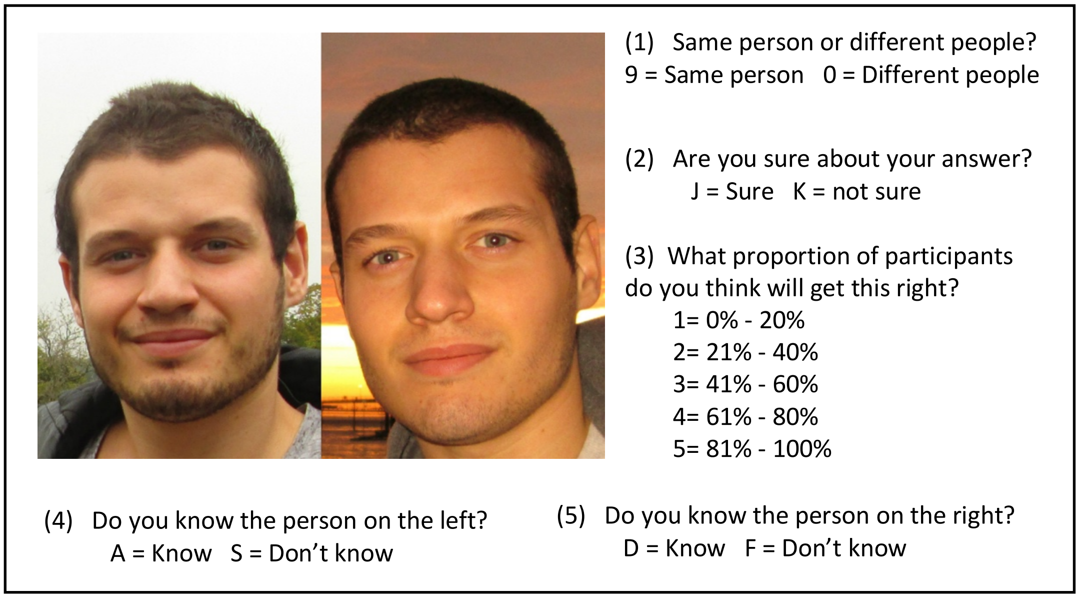
We defined *actual performance* as actual test score, that is, the proportion of correct responses in the matching task. Participants' actual test scores were used to determine their actual percentile ranking (0–100%), and to define performance quartiles for the Dunning–Kruger analyses.

Self-estimates comprised two metrics. *Estimated test score* was an estimate of absolute performance, captured trial-by-trial. Following each identity decision (Same or Different), participants indicated whether they were sure or unsure of their decision. Estimated test score was defined as the number of ‘sure’ responses plus half of the number of ‘unsure’ responses. That is, we assumed that participants guessed on unsure trials and answered half of them correctly by chance. *Estimated percentile ranking* (0–100%) was an estimate of relative performance, reported by each participant at the end of each task.

Peer estimates were also captured trial-by-trial. For each image pair, participants estimated the proportion of respondents who would answer correctly (0–20%, 21–40%, 41–60%, 61–80%, 81–100%). To provide context for these estimates, participants were informed that all respondents were UK students. We note that estimated percentile ranking, reported at the end of each task, combines self-estimate and peer estimate.

Procedure

Each display consisted of a pair of face photographs alongside a set of response options as shown in Figure 2.1.



(1) Same person or different people?
9 = Same person 0 = Different people

(2) Are you sure about your answer?
J = Sure K = not sure

(3) What proportion of participants do you think will get this right?
1= 0% - 20%
2= 21% - 40%
3= 41% - 60%
4= 61% - 80%
5= 81% - 100%

(4) Do you know the person on the left?
A = Know S = Don’t know

(5) Do you know the person on the right?
D = Know F = Don’t know

Figure 2.1 Example identity matching display from Experiment 1. In this example, the two photos show the same unfamiliar face. Participants respond to questions 1–5 for each image pair.

For each pair, participants indicated (i) whether the two photos showed the *Same* person or *Different* people, (ii) whether they were *Sure* or *Unsure* of their decision, (iii) the proportion of participants they thought would give the correct answer, and (iv) whether or not they knew the face in each image. Participants were reminded that they did not need to know the person’s name to know that person’s face. Each display remained on screen until the final response, which immediately initiated the next trial. The experimenter explained the task at the beginning of the session using a printed example display, which showed a face that was not presented in the main experiment. Following this example trial, each participant underwent two blocks of 40 trials each (one Familiar block and one Unfamiliar block). At the end of each block, participants estimated their own performance relative to all participants (percentile ranking) by dragging an onscreen slider (0%, “I think I performed worse than other participants” to 100%, “I think I performed better than other participants”). Participants were able to rest between blocks and initiated the next block by pressing the space bar. The entire test session took approximately 30 minutes to complete.

2.2.2 Results and discussion

Faces in the Familiar condition were familiar to more participants (*M* = 55, *SE* = 1.05) than faces in the Unfamiliar condition (*M* = 5, *SE* = .43) [*t* (158) = 44.51, *p* < .001, *d* = 7.04], confirming that our familiarity manipulation was successful.

In comparing cognition and metacognition, we first examined participants’ insight into their own absolute performance (test score) and relative performance (percentile ranking), by combining actual attainment with self-estimates in the same analyses. Our main focus is the Dunning–Kruger analysis based on performance quartiles. We then examined participants’ insight into other people’s performance, focusing specifically on egocentric bias.

Insight into one’s own performance

Dunning and colleagues (Dunning et al., 2003) established the convention of analysing metacognition data by performance quartiles. In this approach, participants are divided into quartiles according to their actual performance. Estimated performance can then be compared to actual performance in each quartile. Figure 2.2 summarizes this analysis for test score and percentile ranking, separately for the Familiar and Unfamiliar face matching tasks.

Test scores

Familiar face matching

Participants' test scores were submitted to a 2 × 4 mixed ANOVA with the within-subjects factor of Measure (Actual Score, Estimated Score) and the between-subjects factor of Quartile (Lowest, Second, Third, Highest). This analysis revealed a main effect of Measure, with Estimated scores ($M = 96.61$, $SE = .49$) exceeding Actual scores ($M = 95.43$, $SE = .35$) overall [$F\ (1,60) = 5.62$, $p < .05$, $\eta_p^2 = .09$]. Unsurprisingly, there was also a main effect of Quartile, with scores increasing from the lowest quartile to the highest quartile [$F\ (3, 60) = 37.76$, $p < .001$, $\eta_p^2 = .65$]. In keeping with the standard Dunning–Kruger pattern, these main effects were qualified by a significant Measure × Quartile interaction [$F\ (3,60) = 9.58$, $p < .001$, $\eta_p^2 = .32$]. Simple main effects showed that Estimated score exceeded Actual score in the Lowest quartile [$F\ (1,60) = 26.74$, $p < .001$, $\eta_p^2 = .31$], but not in the 2nd [$F\ (1,60) = 1.81$, $p = .18$], 3rd [$F\ (1,60) = .09$, $p = .76$] or Highest quartiles [$F\ (1,60) = 3.28$, $p = .08$]. The simple main effect of Quartile was significant for both Actual scores [$F\ (3,60) = 68.35$, $p < .001$, $\eta_p^2 = .77$] and Estimated scores [$F\ (3,60) = 7.58$, $p < .001$, $\eta_p^2 = .28$].

Unfamiliar face matching

Test scores for the unfamiliar face matching task were analysed in the same way. For unfamiliar faces, there was no difference between Estimated scores ($M = 79.99$, $SE = 1.28$) and Actual scores ($M = 82.05$, $SE = .39$) overall [$F\ (1,60) = 2.83$, $p = .10$]. Again, there was a main effect of Quartile, with scores increasing from the lowest quartile to the highest quartile [$F\ (3, 60) =$

15.69, $p < .001$, $\eta_p^2 = .44$]. There was also a significant crossover interaction between these factors [F (3,60) = 8.42, $p < .001$, $\eta_p^2 = .30$]. Simple main effects showed that Estimated score exceeded Actual score in the Lowest quartile [F (1, 60) = 5.28, $p < .05$, $\eta_p^2 = .08$] but not the 2nd quartile [F (1,60) = .78, $p = .38$]. The effect then reversed in the 3rd [F (1, 60) = 13.02, $p < .01$, $\eta_p^2 = .18$] and highest quartiles [F (1, 60) = 9.73, $p < .01$, $\eta_p^2 = .14$], such that Actual score exceeded Estimated score. The simple main effect of Quartile was significant for Actual scores [F (3, 60) = 133.57, $p < .001$, $\eta_p^2 = .87$], but not for Estimated scores [F (3,60) = 1.53, $p = .22$].

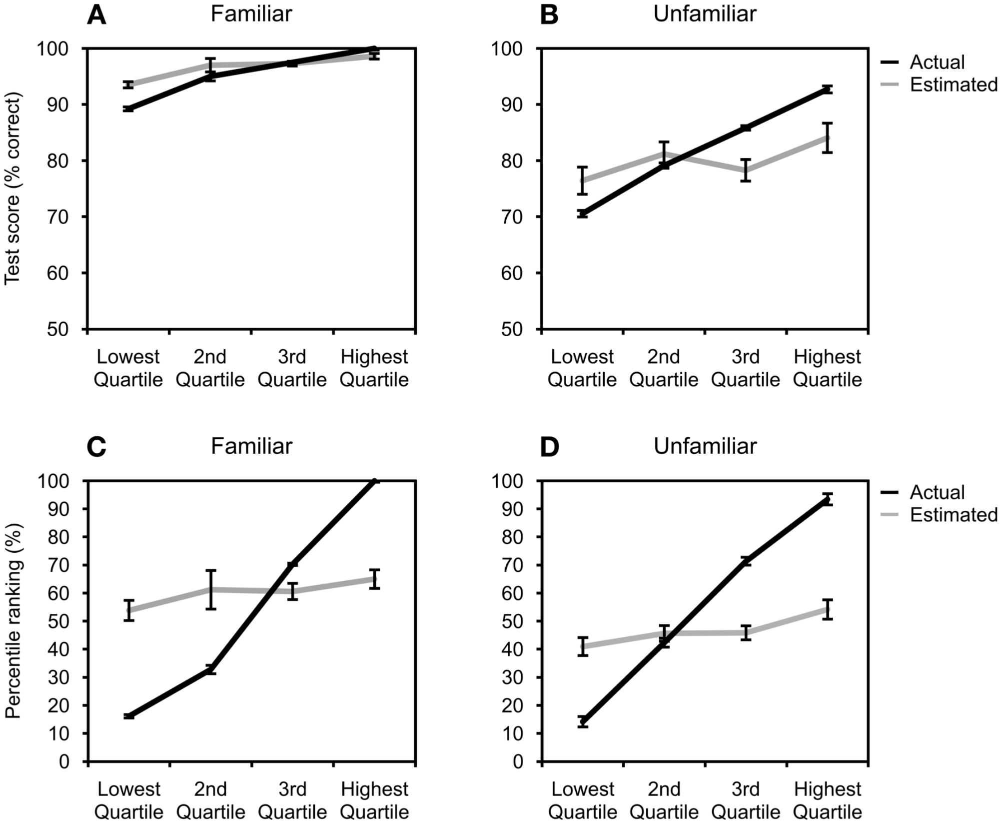


Figure 2.2 Dunning–Kruger analysis of the face matching tasks in Experiment 1. The top row shows test scores for (A) Familiar faces and (B) Unfamiliar faces. Actual scores (black) and Estimated scores (grey) are plotted separately for each performance quartile. Chance performance is 50%. The bottom row shows percentile rankings for (C) Familiar faces and (D)

Unfamiliar faces. Actual ranks (black) and Estimated ranks (grey) are plotted separately for each performance quartile. Error bars show SE.

Percentile ranking

Familiar face matching

As with the test scores, percentile rankings were entered into a 2 × 4 mixed ANOVA with the within-subjects factor of Measure (Actual Score, Estimated Score) and the between-subjects factor of Quartile (Lowest, Second, Third, Highest). This analysis revealed a main effect of Measure, with Estimated rank ($M = 60.15$, $SE = 2.46$) exceeding Actual rank ($M = 54.81$, $SE = .66$) overall [F (1, 60) = 4.65, $p < .05$, $\eta_p^2 = .07$], and the expected main effect of Quartile [F (3, 60) = 91.83, $p < .001$, $\eta_p^2 = .82$]. There was also a significant crossover interaction between Measure and Quartile [F (3, 60) = 62.82, $p < .001$, $\eta_p^2 = .76$]. Simple main effects showed that Estimated rank exceeded Actual rank in the Lowest quartile [F (1, 60) = 82.47, $p < .001$, $\eta_p^2 = .58$] and the 2nd quartile [F (1, 60) = 14.63, $p < .001$, $\eta_p^2 = .20$]. However, this effect was reversed in the 3rd [F (1, 60) = 8.22, $p < .01$, $\eta_p^2 = .12$] and Highest quartiles [F (1, 60) = 84.46, $p < .001$, $\eta_p^2 = .59$], in which Actual rank exceeded Estimated rank. The simple main effect of Quartile was significant for Actual rank [F (3, 60) = 1137.10, $p < .001$, $\eta_p^2 = .98$], but not for Estimated rank [F (3,60) = 1.35, $p = .27$].

Unfamiliar face matching

Percentile ranks for the unfamiliar face task were analysed in the same way. This analysis revealed a significant effect of Measure, with Actual rank ($M = 55.31$, $SE = 1.11$) exceeding

Estimated rank (M= 46.65, SE= 1.75) [F (1, 60) = 17.16, $p < .001$, $\eta_p^2 = .22$], and the expected main effect of Quartile, with ranks increasing from the lowest quartile to the highest quartile [F (3, 60) = 82.98, $p < .001$, $\eta_p^2 = .81$]. As with the familiar face task, there was a significant crossover interaction between Measure and Quartile [F (3, 60) = 46.16, $p < .001$, $\eta_p^2 = .70$]. Simple main effects showed that Estimated rank exceeded Actual rank in the Lowest quartile [F (1, 60) = 37.38, $p < .001$, $\eta_p^2 = .38$] but not the 2nd quartile [F (1,60) = .68, $p = .41$]. This effect was reversed in the 3rd [F (1, 60) = 50.81, $p < .001$, $\eta_p^2 = .46$] and Highest quartiles [F (1, 60) = 68.90, $p < .001$, $\eta_p^2 = .54$], with Actual rank exceeding Estimated rank. The simple main effect of Quartile was significant for Actual rank [F (3, 60) = 219.02, $p < .001$, $\eta_p^2 = .92$], but not for Estimated rank [F (3,60) = 2.06, $p = .12$].

Insight into other people's performance

To assess egocentric bias in each task, we compared peer estimates (attributions of other people's performance) generated by the highest and lowest performing participants. Egocentric bias predicts that peer estimates from the Highest quartile will be higher than peer estimates from the Lowest quartile. Figure 2.3 summarises this analysis separately for the familiar and unfamiliar face matching tasks.

Familiar face matching

For each familiar face, we calculated the mean peer estimate from Lowest quartile and Highest quartile participants. Peer estimates were on a scale of 1–5, where 1 means "0–20% of participants will answer correctly", and 5 means "81–100% of participants will answer correctly" (see Figure 1). An independent t-test confirmed that peer estimates from Highest quartile participants (*M* = 4.67, *SE* = .03) were significantly higher than those from Lowest quartile participants (*M* = 4.37, *SE* = .06) [*t* (78) = 4.72, *p* < .001, *d* = 1.06].

Unfamiliar face matching

Peer estimates in the unfamiliar face matching task were analysed in the same way. An independent t-test showed that peer estimates were higher for the Highest quartile (*M* = 3.81, *SE* = .05) than for the Lowest quartile (*M* = 3.56, *SE* = .05) [*t* (78) = 3.70, *p* < .001, *d* = .83].

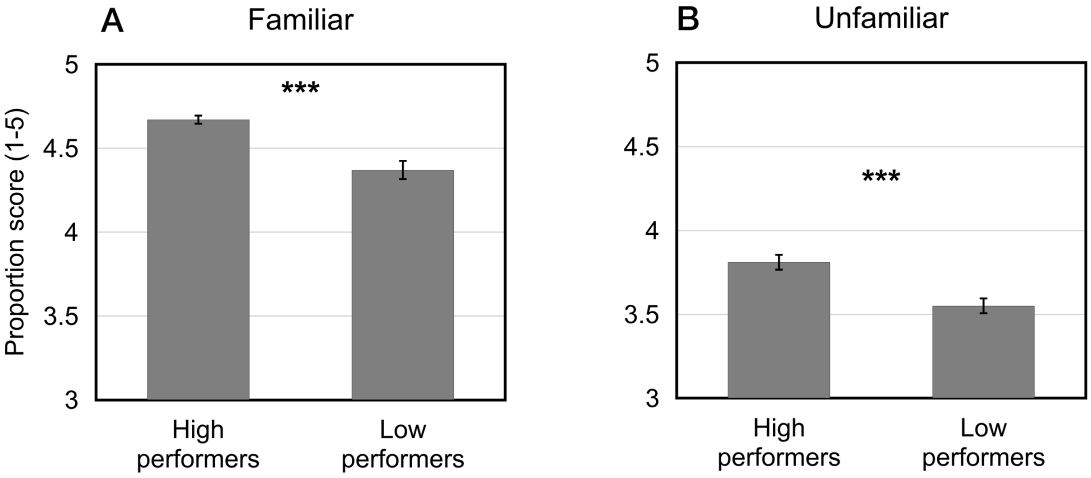


Figure 2.3 Egocentric bias in peer estimates from the face matching tasks in Experiment 1. (A) Familiar faces. (B) Unfamiliar faces. In both tasks, High performers attributed higher performance to others; Low performers attributed lower performance to others. Error bars show SE.

The Dunning–Kruger analysis of test scores (Figure 2.2) showed that self-estimates were higher for familiar faces than for unfamiliar faces. Combining this observation with egocentric bias implies that peer estimates should also be higher for familiar faces than for unfamiliar faces. A within-subjects t-test confirmed that this difference was significant (Familiar $M = 4.59$, $SE = .04$; Unfamiliar $M = 3.69$, $SE = .04$) [$t\ (63) = 19.27$, $p < .001$, $d = 2.41$].

The cognitive aspects of these results were as expected from previous research. When matching faces for identity, accuracy was at ceiling for familiar faces (95% correct overall), and was significantly lower for unfamiliar faces (82% correct overall). We also obtained the expected individual differences in performance. Although unfamiliar face matching was generally poor, some people were much better at it than others (range 60–97.5%). This wide range in performance lends itself to a Dunning–Kruger type of analysis.

Claims of face being 'special' notwithstanding, we found absolutely standard Dunning–Kruger effects in face identification. Low performers overestimated their performance, and high performers underestimate their performance. This pattern emerged in test score (an absolute measure, captured trial by trial), and in percentile rank (a relative measure, captured retrospectively). It also occurred in both Familiar and Unfamiliar identity conditions, though test scores in the Familiar condition were somewhat compressed against ceiling.

We also saw a clear evidence of egocentric bias. High performers made higher peer estimates than low performers; and peer estimates were higher overall for familiar faces than for unfamiliar faces.

All of these findings concern matters of identification. Given that other aspects of face perception (such as gaze direction and emotional expression) are known to dissociate from identification, we next examined metacognition for these other tasks.

2.3 Experiment 2 Identity, gaze, and expression matching

The purpose of our second experiment was to establish whether the metacognitive pattern seen for identification in Experiment 1 extends to other face tasks. Specifically, we asked whether Dunning–Kruger Effects and egocentric bias extend to perception of gaze direction and emotional expression. These tasks are especially interesting from a metacognition perspective. First, gaze direction and emotional expression are dissociable from face identification (Andrews & Ewbank, 2004; Hoffman & Haxby, 2000; Winston, Henson, Fine-Goulden, & Dolan, 2004). This dissociation allows us to test the generalizability of metacognitive patterns across cognitively unrelated tasks. Second, unlike perception of facial identity, perception of gaze direction and emotional expression have both been associated with cognitive insight, in the specific sense of inferring other people's mental states from their behaviour (Calder et al., 2002; Friesen & Kingstone, 1998; Simpson & Crandall, 1972). Given that the ability to infer mental states seems related to metacognition, it is possible that individuals who perform especially well in these tasks will also demonstrate especially high metacognitive insight (and vice versa).

To extend our analysis to 'cognitive insight' signals from the face, we adapted the identity matching task from Experiment 1 to assess perception of gaze direction and emotional expression. To allow replication of key findings, and to facilitate comparison across diverse tasks, we also repeated the unfamiliar face matching task from Experiment 1. The task format (Same/Different judgements to paired images) and task measures (Actual versus Estimated test scores and percentile ranks) were the same in all three tasks. This homology ensured that data from all three tasks could be analysed in the same way.

Based on previous studies, we expected actual performance on the identity, gaze, and expression tasks to be either uncorrelated (identity versus gaze; identity versus expression) or weakly correlated (gaze versus expression). Our main interest was whether similar metacognitive patterns emerged in all three tasks. If our gaze and expression tasks require metacognitive insight, then people with the greatest insight should perform best, and people with the least insight should perform worst. In that case, the Dunning–Kruger Effect and the egocentric bias should break down. On the other hand, if metacognitive biases generalize even across tasks that are not correlated at the cognitive level, then the Dunning–Kruger Effect and the egocentric bias should persist in all three tasks.

2.3.1 Method

Participants

Sixty-four UK students (56 female, 8 male; mean age = 20 years; age range 18–26 years) from the University of York took part in exchange for a small payment or course credit. None of these volunteers participated in Experiment 1.

Stimuli

Face identity task

Stimuli for the identity matching task were the same as for the unfamiliar face matching task in Experiment 1 (See Figure 1 and Figure 4). As all of the faces were now unfamiliar, we omitted the image-by-image familiarity check (Questions 4 & 5 in Figure 1).

Gaze direction task

Stimuli for the gaze matching task were drawn from Jenkins, Beaver, and Calder (2006). We selected eight models (4 female, 4 male), each posing five gaze directions (10° left [L10], 5° left [L05], straight ahead [S00], 5° right [R05], 10° right [R10]; 40 images in total). Each face was presented in an elliptical mask measuring 230 pixels high × 205 pixels wide. Stimulus pairs always combined two identities of the same sex. For each combination, we created a Same Direction pair (two faces looking in the same direction: L10, L05, S00, R05, or R10) and a Different Direction pair (two faces looking in different directions). To ensure a range of difficulty, Different Direction pairs differed by 5° (S00 vs R05; S00 vs R05), 10° (L05 vs R05), or 20° (L10 vs R10; R10 vs L10). To make deviations from the midline easier to discern, the two faces in each pair were arranged vertically rather than horizontally (see Figure 2.4). Each face appeared once at the top and once at the bottom in both a Same Direction and a Different Direction trial, resulting in a total of 80 trials.

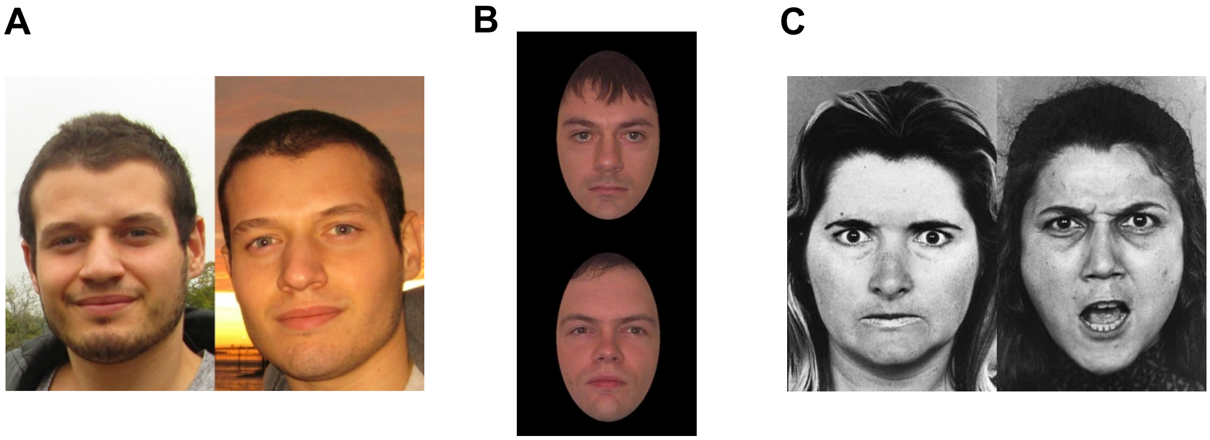


Figure 2.4 Example face matching stimuli from Experiment 2. (A) Identity matching. (B) Gaze matching. (C) Expression matching.

Expression task

Stimuli for the expression matching task were drawn from the Facial Expressions of Emotion: Stimuli and Tests (FEEST) dataset (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). Given that facial expressions of happiness are reliably recognized (Ekman, Friesen, & Ellsworth, 1972; Calvo & Lundqvist, 2008), we excluded happiness images to avoid ceiling effects. We selected five female models, each posing five facial expressions of emotion (anger, disgust, fear, surprise, and sadness; 25 images in total). Each face image measured 362 pixels high × 241 pixels wide. Stimulus pairs always combined two identities. Each image was combined with each identity in a Same Emotion pair (two faces expressing the same emotion) and a Different Emotion pair (two faces expressing different emotions), resulting in a total of 100 trials. The two images in each pair were arranged horizontally (see Figure 4). Each identity and each emotion appeared equally often on the left and on the right.

Design

All participants completed the *Identity*, *Gaze*, and *Expression* matching tasks in separate blocks. Block order was counterbalanced with respect to participants so that each task could be encountered first, second, or third. Within each block, trials were presented in a random order. All participants contributed the same measures in all three tasks—actual performance, self-estimates, and peer estimates.

Procedure

The procedure was the same as in Experiment 1 except for the following changes. Participants now completed three matching tasks (*Identity*, *Gaze Direction*, *Emotional Expression*), making *Same*/*Different* judgements according to the task. As before, participants indicated whether they were *Sure* or *Unsure* of each decision, and estimated the proportion of participants (UK students) they thought would give the correct answer. The entire test session took approximately 40 minutes to complete.

2.3.2 Results and discussion

Before proceeding to the metacognitive analyses, we first examined performance on each of the three face matching tasks. At the group level, actual scores were very similar for the three tasks (Identity M= 78.91, SE = .39; Gaze M= 80.54, SE = .41; Expression M= 82.21, SE = .30), indicating similar levels of overall difficulty. Importantly however, there was no significant correlation between actual scores in the Identity and Gaze tasks [r (62) = .13, p = .31], or between the Identity and Expression tasks [*r* (62) = .15, *p* = .25]. There was a moderate correlation between actual scores in the Gaze and Expression tasks [*r* (62) = .30, *p* < .05]. For actual rankings, there were no significant correlations between any of the tasks [Identity and Gaze *r* (62) = .15, *p* = .25; Identity and Expression *r* (62) = .20, *p* = .12; Gaze and Expression *r* (62) = .17, *p* = .17]. In sum, the pattern of performance is as expected based on previous work. Invariant and changeable aspects of faces cleave together to some extent, but correlations between different face tasks are otherwise low.

Our metacognitive analysis follows the same plan as Experiment 1. We first examine participants' insight into their own absolute performance (test score) and relative performance (percentile ranking), by combining actual attainment with self-estimates in a Dunning–Kruger analysis for each task. We then examine participants' insight into other people's performance, focusing specifically on egocentric bias. Finally, we consider the stability of cognition and metacognition across different face tasks.

Insight into one's own performance

As in Experiment 1, participants were divided into quartiles according to actual performance. Estimated performance was then compared to actual performance in each quartile. Figure 2.5 summarises this analysis for test score and percentile ranking, separately for the Identity, Gaze, and Expression matching tasks.

Test scores

Identity matching

Test scores were submitted to a 2 × 4 mixed ANOVA with the within-subjects factor of Measure (Actual Score, Estimated Score) and the between-subjects factor of Quartile (Lowest, Second, Third, Highest). This analysis revealed a main effect of Measure, with Estimated scores (M = 83.03, SE = 1.04) exceeding Actual scores (M= 78.91, SE = .39) overall [F (1, 60) = 14.25, p < .001, η_p^2 = .19]. There was also a main effect of Quartile, with scores increasing from the lowest quartile to the highest quartile [F (3, 60) = 29.28, p < .001, η_p^2 = .59]. These main effects were qualified by a significant Measure × Quartile interaction [F (3, 60) = 19.46, p < .001, η_p^2 = .49]. Simple main effects showed that Estimated score exceeded Actual score in the Lowest quartile [F (1,60) = 59.69, p < .001, η_p^2 = .50] and the 2nd quartile [F (1,60) = 6.73, p < .05, η_p^2 = .10], but not for the 3rd quartile [F (1,60) = .12, p = .73]. The effect then reversed in the Highest quartiles [F (1,60) = 8.14, p < .01, η_p^2 = .12]. The simple main effect of Quartile was significant for Actual scores [F (3,60) = 192.92, p < .001, η_p^2 = .91] but not for Estimated scores [F (3,60) = .61, p = .61].

Gaze matching

Test scores for the gaze matching task were analysed in the same way. Again, there was a main effect of Measure, with Estimated scores (M = 91.59, SE = .76) exceeding Actual scores (M= 80.54, SE = .41) overall [F (1, 60) = 166.87, p < .001, η_p^2 = .74] and a main effect of Quartile, with scores increasing from the lowest quartile to the highest quartile [F (3, 60) = 20.03, p < .001, η_p^2 = .50]. These main effects were also qualified by a significant Measure × Quartile interaction [F (3, 60) = 17.48, p < .001, η_p^2 = .47]. Simple main effects showed that Estimated score exceeded Actual score in the Lowest quartile [F (1,60) = 145.31, p < .001, η_p^2 = .71], the 2nd quartile [F (1,60) = 64.21, $p < .001$, $\eta_p^2 = .52$] and the 3rd quartile [F (1,60) = 20.64, $p < .001$, $\eta_p^2 = .26$], but not for the Highest quartiles [F (1,60) = 3.21, $p = .08$]. The simple main effect of Quartile was significant for Actual scores [F (3,60) = 83.37, $p < .001$, $\eta_p^2 = .81$] but not for Estimated scores [F (3,60) = .56, $p = .65$].

Expression matching

For the Expression task, there was a main effect of Measure, with Estimated scores ($M = 89.69$, $SE = .83$) exceeding Actual scores ($M = 82.21$, $SE = .30$) overall [F (1,60) = 75.33, $p < .001$, $\eta_p^2 = .56$] and a main effect of Quartile, with scores increasing from the lowest quartile to the highest quartile [F (3,60) = 9.67, $p < .001$, $\eta_p^2 = .33$]. These main effects were qualified by a significant Measure × Quartile interaction [F (3, 60) = 16.97, $p < .001$, $\eta_p^2 = .46$]. Simple main effects showed that Estimated score exceeded Actual score in the Lowest quartile [F (1,60) = 94.51, $p < .001$, $\eta_p^2 = .61$] and the 2nd quartile [F (1,60) = 37.12, $p < .001$, $\eta_p^2 = .38$], but not for the 3rd quartile [F (1,60) = 2.62, $p = .11$] or the Highest quartiles [F (1,60) = .17, $p = .68$]. The simple main effect of Quartile was significant for Actual scores [F (3,60) = 112.10, $p < .001$, $\eta_p^2 = .85$] but not for Estimated scores [F (3,60) = .50, $p = .69$].

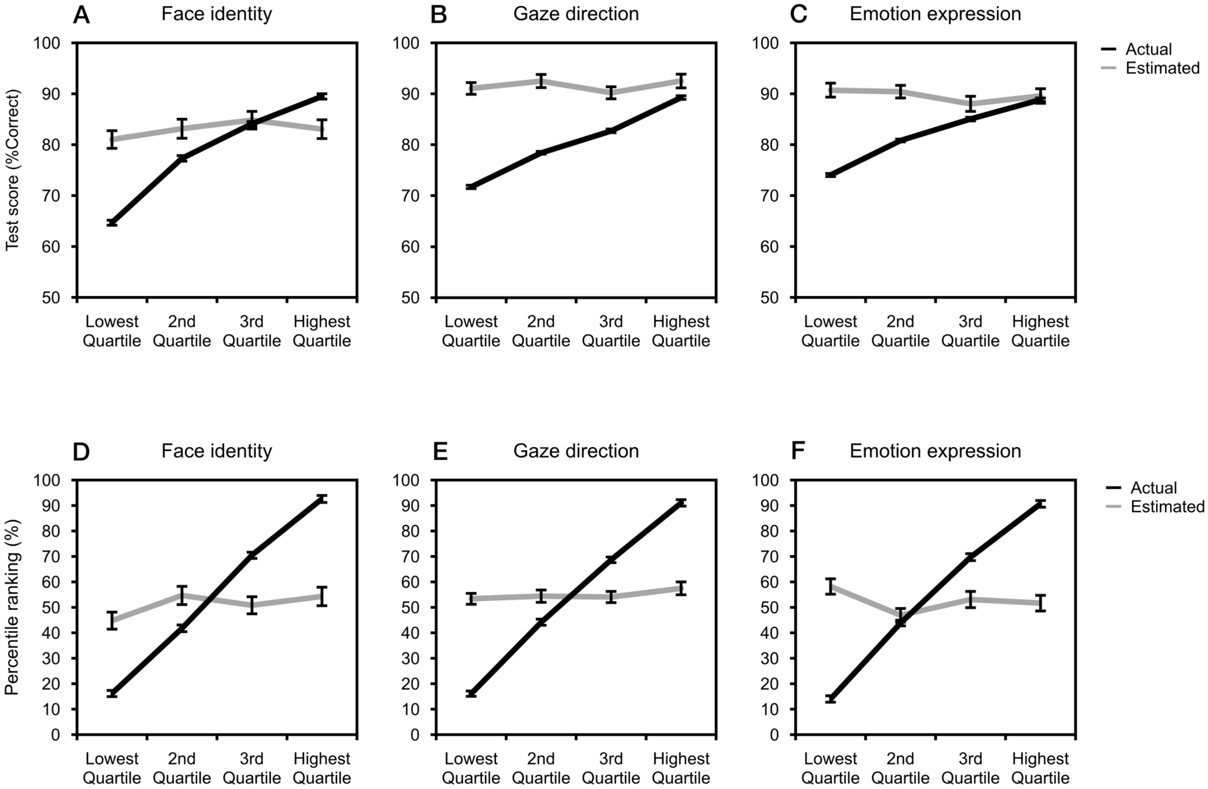


Figure 2.5 Dunning–Kruger analysis of the face matching tasks in Experiment 2. The top row shows test scores for (A) Identity, (B) Gaze, and (C) Expression. Actual scores (black) and Estimated scores (grey) are plotted separately for each performance quartile. Chance performance is 50%. The bottom row shows percentile rankings for (D) Identity, (E) Gaze, and (F) Expression. Actual ranks (black) and Estimated ranks (grey) are plotted separately for each performance quartile. Error bars show SE.

Percentile ranking

Identity matching

As with the test scores, percentile rankings were entered into a 2 × 4 mixed ANOVA with the within-subjects factor of Measure (Actual Score, Estimated Score) and the between-subjects factor of Quartile (Lowest, Second, Third, Highest). On this occasion, the overall difference between Estimated rank (*M* = 51.19, *SE* = 2.01) and Actual rank (*M* = 55.29, *SE* = .91) was not significant [F (1, 60) = 3.58, $p = .06$]. There was the expected main effect of Quartile [F (3, 60) = 67.01, $p < .001$, $\eta_p^2 = .77$] and a significant crossover interaction between Measure and Quartile [F (3, 60) = 49.41, $p < .001$, $\eta_p^2 = .71$]. Simple main effects showed that Estimated rank exceeded Actual rank in the Lowest quartile [F (1, 60) = 46.70, $p < .001$, $\eta_p^2 = .44$] and the 2nd quartile [F (1, 60) = 8.44, $p < .01$, $\eta_p^2 = .12$]. However, this effect was reversed in the 3rd [F (1, 60) = 22.01, $p < .001$, $\eta_p^2 = .27$] and Highest quartiles [F (1, 60) = 73.78, $p < .001$, $\eta_p^2 = .55$], in which Actual rank exceeded Estimated rank. The simple main effect of Quartile was significant for Actual rank [F (3, 60) = 334.93, $p < .001$, $\eta_p^2 = .94$], but not for Estimated rank [F (3,60) = 1.33, $p = .27$].

Gaze matching

Again, there was no overall difference between Estimated rank (M = 54.88, SE = 1.44) and Actual rank (M = 55.05, SE = .85) [F (1, 60) = .01, $p = .91$]. There was a main effect of Quartile [F (3, 60) = 85.36, $p < .001$, $\eta_p^2 = .81$] and a significant crossover interaction between Measure and Quartile [F (3, 60) = 109.75, $p < .001$, $\eta_p^2 = .85$]. Simple main effects showed that Estimated rank exceeded Actual rank in the Lowest quartile [F (1, 60) = 179.62, $p < .001$, $\eta_p^2 = .75$] and the 2nd quartile [F (1, 60) = 11.34, $p < .01$, $\eta_p^2 = .16$]. However, this effect was reversed in the 3rd [F (1, 60) = 26.15, $p < .001$, $\eta_p^2 = .30$] and Highest quartiles [F (1, 60) = 113.40, $p < .001$, $\eta_p^2 = .65$], in which Actual rank exceeded Estimated rank. The simple main effect of Quartile was significant for Actual rank [F (3, 60) = 364.52, $p < .001$, $\eta_p^2 = .95$], but not for Estimated rank [F (3,60) = .37, $p = .78$].

Expression matching

As with the other tasks, there was no overall difference between Estimated rank (*M* = 52.52, *SE* = 1.78) and Actual rank (*M* = 54.62, *SE* = .92) [*F* (1, 60) = 1.19, *p* = .28]. Again, the results showed the expected main effect of Quartile [*F* (3, 60) = 57.28, $p < .001$, $\eta_p^2 = .74$] and a significant crossover interaction between Measure and Quartile [*F* (3, 60) = 82.56, $p < .001$, $\eta_p^2 = .81$]. Simple main effects showed that Estimated rank exceeded Actual rank in the Lowest quartile [*F* (1, 60) = 133.19, $p < .001$, $\eta_p^2 = .69$] but not for the 2nd quartile [*F* (1, 60) = .75, *p* = .39]. This effect was reversed in the 3rd [*F* (1, 60) = 16.50, $p < .001$, $\eta_p^2 = .22$] and Highest quartiles [*F* (1, 60) = 97.43, $p < .001$, $\eta_p^2 = .62$], in which Actual rank exceeded Estimated rank. The simple main effect of Quartile was significant for Actual rank [*F* (3, 60) = 322.44, $p < .001$, $\eta_p^2 = .94$], but not for Estimated rank [*F* (3,60) = 1.88, *p* = .14].

Insight into other people's performance

Peer estimates in the three tasks were analysed in the same way as in Experiment 1. Figure 2.6 summarises the results of this analysis.

Identity matching

Despite the numerical difference, peer estimates from Highest quartile participants (*M* = 4.13, *SE* = .05) were not significantly higher than those from Lowest quartile participants (*M* = 4.03, *SE* = .03) [*t* (78) = 1.82, *p* = .07, *d* = .41].

Gaze matching

As expected, peer estimates from Highest quartile participants (*M* = 4.45, *SE* = .04) were significantly higher than those from Lowest quartile participants (*M* = 4.23, *SE* = .03) [*t* (158) = 4.42, *p* < .001, *d* = .70].

Expression matching

Here too, peer estimates from Highest quartile participants (*M* = 4.46, *SE* = .03) were significantly higher than those from Lowest quartile participants (*M* = 4.28, *SE* = .02) [*t* (198) = 5.00, *p* < .001, *d* = .71].

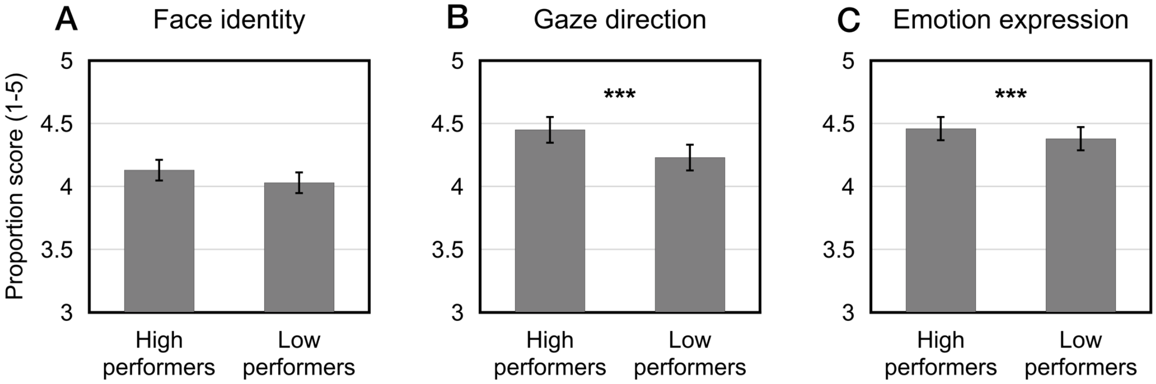


Figure 2.6 Egocentric bias in peer estimates from the (A) Identity, (B) Gaze, and (C) Expression matching tasks in Experiment 2. High performers attributed higher performance to others; Low performers attributed lower performance to others. Error bars show SE.

One interesting aspect of these findings concerns the Dunning–Kruger analysis of test scores (Figure 2.5). For high performers in the Gaze and Expression tasks, Actual Scores and Estimated scores converged, but did not cross over. On its own, this pattern may appear to support the idea that gaze and expression perception and metacognition have some shared basis: those who performed best on these matching tasks also showed the most accurate insight into their performance. However, two observations caution against this interpretation. First, high performers did not show accurate insight in the relative measure (percentile ranking; Figure 2.5), or when estimating the performance of others (Figure 2.6). Second, estimated scores in the Gaze and Expression tasks were high for all performance quartiles. Why should people think they are so good at these particular tasks? One possibility is poor calibration. Everyday life might provide less useful feedback on errors of gaze and expression (which can vary continuously) compared with errors of identity (which varies discretely). If that is the case, then people should have less insight into their fallibility in gaze and expression tasks. One way to test this possibility is through feedback training. If people receive feedback on their gaze and expression perception, their self-estimates should fall accordingly.

Stability of cognition and metacognition across face tasks

The preceding analyses show that Dunning–Kruger effects arise in a range of different face tasks. In all of these tasks, low performers overestimated their performance. For high performers, this tendency was reversed or eliminated. Multiple measures of performance give us the opportunity to examine the stability of Dunning–Kruger effects across tasks. Do people who overestimate themselves in one task also overestimate themselves in the other tasks? Or is assessment of one's own performance (like performance itself) task dependent? Figure 2.7 shows the stability of performance across tasks.

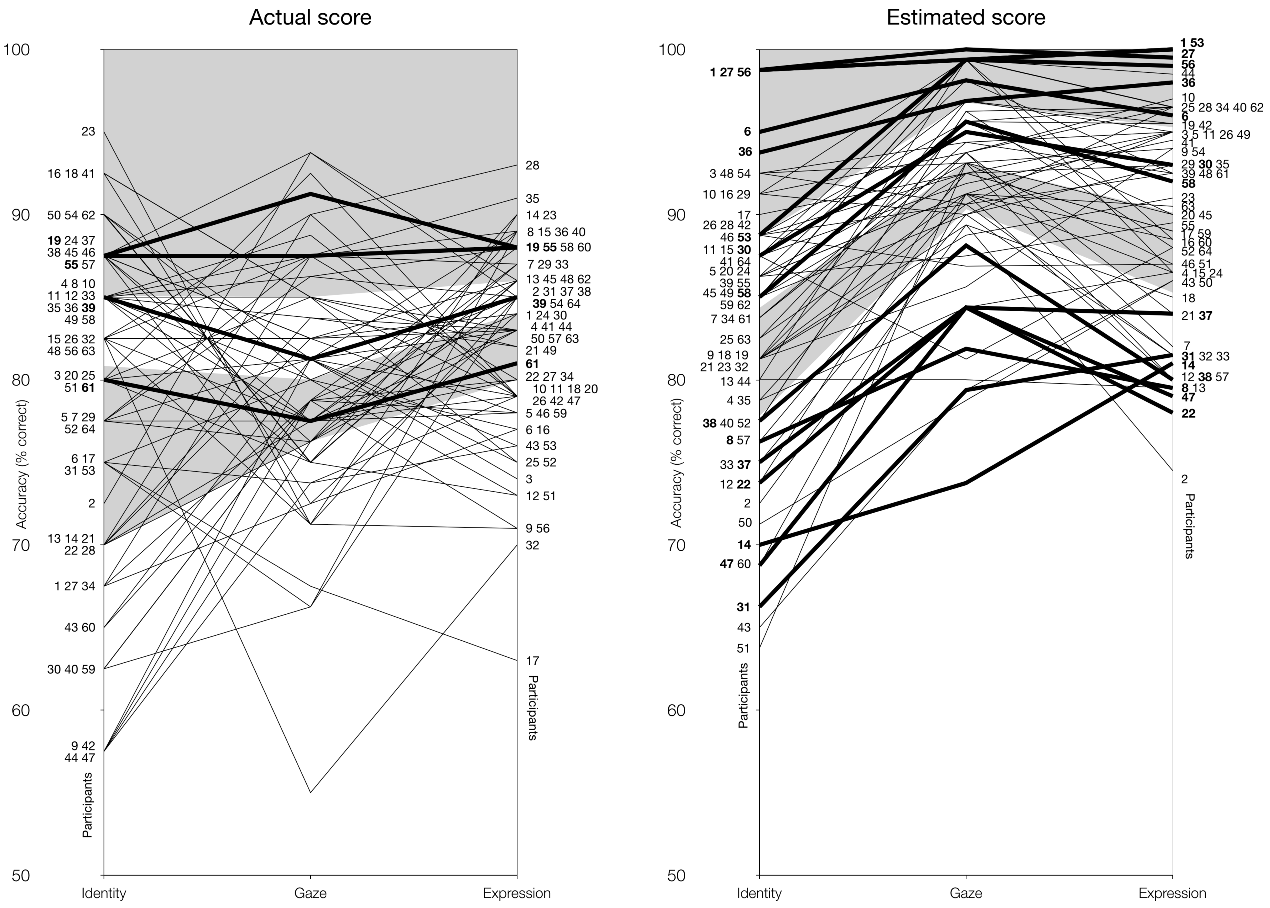


Figure 2.7 Stability of performance across the three face matching tasks in Experiment 2. Actual scores are shown on the left. Estimated scores are shown on the right. Grey and white regions in each panel are performance quartiles. Heavy lines indicate participants who stayed within the same performance quartile across all three tasks. Light lines indicate participants who switched between performance quartiles.

For Actual Scores, only 4 participants stayed within the same performance quartile across all three tasks. For Estimated Scores, 15 participants stayed within the same quartile. In other words, self-assessment was more stable than ability [χ^2 (1) = 5.26, $p < .05$; OR = 3.75, 95% CI 1.24–11.30]. This pattern suggests that the tendency to overestimate or underestimate one's own performance is not strictly task dependent. We return to this issue in the General Discussion section.

2.4 General Discussion

Unusually for studies of face perception, the experiments reported here concern (i) the cognitive level, (ii) the metacognitive level, and (iii) the relation between these two levels. We first summarize the findings for each of these areas in the context of previous research, before moving on to theoretical and applied implications.

At the cognitive level, performance on the individual face tasks was as expected from previous findings. In the identity matching task, overall accuracy was lower for unfamiliar faces (82% in Experiment 1; 79% in Experiment 2) than for familiar faces (96%), demonstrating the standard familiarity advantage (e.g. Burton, White, & McNeill, 2010; Noyes & Jenkins, 2017, 2019). Paired matching has not been widely used to assess gaze perception or processing of emotional expression, but the present findings demonstrate the applicability of this method to both tasks. Overall accuracy rates were similar across unfamiliar identity, gaze direction, and emotional expression tasks (~80%), and within each task, the range of scores (~55–95%) allowed meaningful analysis of individual differences. Critically, this analysis revealed little or no correlation among scores on the three tasks. That is, a person's score on one task tells us very little about their scores on the other two tasks, even though all three tasks concern face perception. The observed dissociations among these scores are consistent with previous behavioural and neural evidence for independence among face perception abilities (e.g. Young, Newcombe, de Haan, Small, & Hay, 1993; Duchaine, Jenkins, Germine, & Calder, 2009). However, previous studies have used different tasks, different measures, and different groups to gauge face perception abilities. This is the first time that three such abilities have been assessed in a within-subjects design, using the common task of paired matching. One advantage of this approach is that it imposed the same level for chance performance in all three tasks (50%). This uniformity facilitates comparisons across tasks. It also provides a stable baseline against which to compare metacognitive judgements of one's own and other people's ability.

At the metacognitive level, our findings concern to two processes—*self*-estimates (insight into one's own cognition) and *peer* estimates (insight into other people's cognition). Our self-estimate measures extend Kruger & Dunning's (1999) 'unskilled and unaware' effect into the novel domain of face perception. In identity matching for both familiar faces (Experiment 1) and unfamiliar faces (Experiments 1 and 2), low performers overestimated their own absolute accuracy (percent correct score), and high performers underestimated their own absolute accuracy, giving rise to a classic crossover interaction between estimated test score and actual test score. In matching for gaze direction and for emotional expression (Experiment 2), estimated accuracy levels were higher overall than for the identity tasks. Thus, while low performers again overestimated their own accuracy, for high performers this tendency was merely eliminated rather than being reversed as it was in the identity tasks. For relative accuracy (rank), the picture was clear cut. In all four matching tasks (Experiments 1 and 2), low performers overestimated their rank, and high performers underestimated their rank. These measures are consistent in showing that participants had rather little insight into their own ability—neither their absolute accuracy level, nor their standing in relation to other people.

Our peer estimate measures also showed a consistent pattern. In identity matching for familiar faces (Experiment 1) and unfamiliar faces (Experiments 1 and 2), high performers attributed higher accuracy to other people than did low performers. Similar performance-contingent estimates emerged in the gaze direction and emotional expression tasks (Experiment 2). One possible interpretation of these performance-contingent effects is that participants estimated other people's ability from their own perspective—that is, with an egocentric bias (Ritchie et al., 2015). On this account, high performers presumed that others can do what they themselves can do, while low performers presumed that others cannot do what they themselves cannot do. The metacognitive picture can be summed up as follows. People estimated their own face perception performance with an "unskilled and unaware" bias, and estimated other people's performance with an egocentric bias.

One interesting aspect of our findings is the consistency of *Estimated* performance across tasks, relative to *Actual* performance across tasks. This pattern suggests that self-estimates are not driven solely by insight into one's own performance, but also involve some determinant that is more stable across tasks. Although the current data do not allow us to single out specific determinants, individual differences in general intelligence or personality could play a role. On a personality account, some participants tend to imagine that they are doing rather well, irrespective of the task, while others tend to imagine that they are doing rather poorly, irrespective of the task. Several previous studies have reported effects of personality traits on self-estimates in other cognitive domains outside of face perception (e.g. narcissism, Ames & Kammrath, 2004; Big Five, Soh & Jacobs, 2013). Combining personality measures with face perception tasks could help to explain the stability of self-assessments seen here. One interesting question is whether the same personality traits predict self-estimates across domains, or whether any domain-specificity emerges. For example, narcissism might inflate self-assessments generally, whereas extroversion might disproportionately inflate self-assessment of socially relevant abilities, such as face perception. Combined testing should distinguish these possibilities.

As well as their theoretical interest, our findings have implications for face perception in clinical and forensic settings. Several clinical disorders are characterised by specific face perception deficits. In this context, unreliability of self-estimates could influence engagement with clinical services. People with developmental prosopagnosia often have little insight into their own impaired facial identification (Fine, 2012). People with autism spectrum disorders (ASD) may not be aware that they have trouble reading social signals from faces (see Bishop & Seltzer, 2012; Schriber, Robins, & Solomon, 2014, for discussions of self-insight in ASD). If people do not realise that their ability is outside the normal range, they may not seek appropriate help (Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008).

Unreliability of peer estimates also has practical implications. There is some evidence that people attribute above-average face recognition ability to individuals with professional training and experience. For example, participants in Ritchie et al.'s (2015) study predicted that passport officers would outperform students at unfamiliar face matching. In fact, training and experience have no appreciable impact on face recognition ability (Towler et al., 2019; White, Kemp, Jenkins, Matheson, & Burton, 2014). Specialists are generally indistinguishable from university students in terms of task performance (Burton, Wilson, Cowan, & Bruce, 1999; White, Kemp, Jenkins, Matheson, & Burton, 2014). A dissociation between estimated and actual performance of specialists could help to explain the enduring popularity of photo-ID as a means of identifying people, despite evidence of its unreliability (Ritchie et al., 2015).

In future work, it would be interesting to compare estimated and actual performance of automatic face recognition systems on face perception tasks. Although there is a huge literature on automatic face recognition (Ranjan et al., 2018; Phillips et al., 2018), very little is known about human understanding of its accuracy. For now, we show that Dunning–Kruger effects and egocentric bias both arise in face perception. Our findings urge caution when interpreting self-report measures of face perception ability. They also reveal a fundamental source of uncertainty in social interactions.

Chapter 3 Familiarity matters more than race in face recognition memory

Reference:

Zhou, X., A.M. Burton & Jenkins, R. (2020). Familiarity matters more than race in face recognition memory. Manuscript submitted for publication.

Abstract

Concerns about face recognition often focus on the other-race effect (ORE), the phenomenon whereby own-race faces are better remembered than other-race faces. Despite this focus, previous studies have not put the magnitude of ORE in context. Here we compared the effects of (i) a race manipulation (own-race/other-race face) and (ii) a familiarity manipulation (familiar/unfamiliar face) in a 2 × 2 factorial design. We found that the familiarity effect was several times larger than the race effect in all performance measures. However, participants expected race to have a larger effect on others than it actually did. Face recognition accuracy depends much more on whether you know the person's face than whether you share the same race. The focus on the other-race effect, in the scientific literature and beyond, is out of proportion with its cognitive importance.

3.1 Introduction

What makes a face hard to recognise? There is a common understanding among scientists, policy makers, and the general public that other-race faces are harder to recognise than own-race faces—a phenomenon known as the other-race effect (ORE) (Meissner & Brigham, 2001). However, the importance of another critical factor, the viewer's familiarity with a face (Burton, Jenkins, & Schweinberger, 2011; Natu & O'Toole, 2011), has not had such broad impact (Table 3.1). From a performance standpoint, the relative attention these effects have received is surprising, as it seems to invert their relative potency. Effects of familiarity on face recognition are generally large (Bruce, 1982; Burton, Wilson, Cowan, & Bruce, 1999), whereas effects of race on face recognition are generally small (e.g. Tanaka, Kiefer, & Bukach, 2004; Sangrigoli et al., 2005).

One possible explanation for this inversion relates to the social importance of racial equality. This could raise the salience of experimental findings that are modulated by race, and perhaps increase the expectation that such effects will be large. The accuracy of metacognitive insight into performance is key here (Zhou & Jenkins, 2020). However, previous studies have rarely examined ORE and face familiarity in the same experiment or incorporated metacognitive measures. As a result, the relative magnitude of these effects has not been a natural comparison.

Table 3.1 **Prominence of race and familiarity effects.** Google search results (number of hits) for race-related and familiarity-related face recognition searches in general and specific sources. For all searches, race hits outnumber familiarity hits. Similar comparisons yield similar results. Searches conducted June 2020.

Search term	Google	Google Scholar	Google Scholar all in title
race face recognition	136,000,000	2,670,000	197
familiarity face recognition	26,400,000	488,000	67
race face recognition *-familiarity*	98,600,000	2,460,000	196
familiarity face recognition *-race*	6,820,000	326,000	66

Here, we manipulated *race* and *familiarity* simultaneously in the same Old/New recognition memory experiments (Bower & Karlin, 1974; O'Toole, Deffenbacher, Valentin, & Abdi, 1994). To gauge metacognitive insight into the effects of these factors, we asked participants to estimate their own performance on this task (self-estimates), and to estimate the performance of other participants (peer-estimates). We expected that familiarity would be a stronger determinant of face recognition memory than race. We also expected that participants would overestimate effects of race relative to familiarity.

3.2 Results

The results of Experiment 3 (N = 60; figure 3.1a–c, figure 3.2) show that recognition performance (*d'*) was affected by *Race* [F (1, 59) = 26.55, p < .001, η^2 = .03], but was much more strongly affected by *Familiarity* [F (1, 59) = 128.57, p < .001, η^2 = .33]. This disparity was present in signal detection analysis, overall accuracy, and statistical effect size, and was not explained by differences in encoding effort during the learning phase (figure 3.2l). In

Experiment 4 (N = 60), a different photo of each face was used at learning and test, hence eliminating image specific memory and requiring person memory (Bruce, 1982; Jenkins, White, Van Montfort, & Burton, 2011). This alteration increased the dominance of *Familiarity* [F (1, 59) = 343.08, $p < .001$, $\eta^2 = .60$] over *Race* [F (1, 59) = 35.51, $p < .001$, $\eta^2 = .02$] (figure 3.1d–f, figure 3.3), demonstrating that the results in Experiment 3 were not due to image repetition.

Self-estimates showed that participants were aware that familiarity affected their own performance more than race (figure 3.2d–k, figure 3.3d–k). In both experiments, trial-by-trial confidence ratings from the learning phase and the test phase tracked actual performance, as did retrospective estimates at debrief. Peer estimates underestimated other people's performance in every condition, consistent with previous work (Alicke et al., 1995). Critically however, they overestimated ORE, giving as much weight to race effects as to familiarity effects (figure 3.2m–o, figure 3.3m–o).

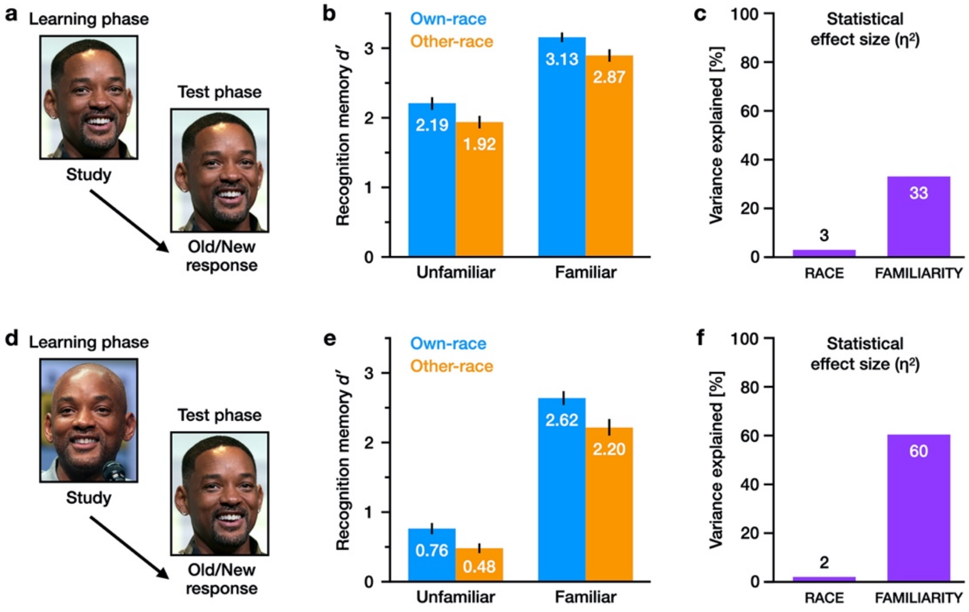


Figure 3.1 **Summary methods and results.** (a) In Experiment 3, each identity was represented by a single image, which was repeated in the learning phase and the test phase (figure 3.2a,d). This standard method is used in most previous studies, but is known to conflate face recognition with image recognition. (b) Mean recognition memory performance (*d'*) in each condition, and (c), statistical effect sizes (η^2) for *Race* and *Familiarity* in Experiment 3. (d) In Experiment 4, each identity was represented by different images in the learning phase and the test phase (figure 3.3a,d). This stronger test minimizes any influence of image recognition, and more closely captures face recognition in the real world. (e) Mean recognition memory performance (*d'*) in each condition, and (f), statistical effect sizes (η^2) for *Race* and *Familiarity* in Experiment 4. Error bars show standard error (SE). [photos by Gage Skidmore CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:Will_Smith_by_Gage_Skidmore.jpg;https://commons.wikimedia.org/wiki/File:Will_Smith_by_Gage_Skidmore_2.jpg].

3.3 Discussion

Directly comparing ORE and familiarity effects puts their relative magnitudes in context. All of our analyses revealed at least a three-fold dominance of familiarity over race in determining recognition performance. This outcome does not diminish the practical importance of ORE for situations where familiarity is constrained (White et al., 2014; Jenkins, Dowsett, & Burton, 2018). However, in many security and forensic settings (e.g. eyewitness testimony; Lindsay et al., 2011), familiarity can vary dramatically. In that situation, reliability of memory is determined more strongly by knowing the person’s face than by sharing the person’s race. Although observers were attuned to the impact of both factors on their own performance, they expected ORE to affect other people more than it did. Over-attribution of ORE, together with the visual salience of race (Valentine, 1991) could help to explain the endurance of ORE as a research topic. It could also explain the recent focus on racial bias in face recognition algorithms

(Cavazos, Phillips, Castillo, & O'Toole, 2019). While familiarity is more important than race in human face recognition, it does not have a clear analogue in machine systems (Jenkins & Burton, 2008). We conclude that the relative prominence of other-race effects and familiarity effects inverts their cognitive importance. Widespread awareness of race effects may detract from scientific focus on more fundamental processes.

3.4 Methods and analysis

3.4.1 Experiment 3 (Same image at learning and test)

Participants

Sixty-four UK students (32 black, 32 white; 48 female, 16 male; mean age 26 years; age range 18–61 years) from the University of York took part in the experiment in exchange for a small payment or course credit. The experiments in this study were approved by the Ethics Committee at the University of York. All participants provided written informed consent.

Stimuli and Design

Full colour face photographs of 32 black celebrities, 32 white celebrities, 32 black non-celebrities, and 32 white non-celebrities were downloaded from online sources (128 faces in total). Each of these categories contained 16 females and 16 males. The names of the celebrities are listed at Appendix A. For each celebrity, we sought a non-celebrity whose face matched the same basic description. Each image was cropped and resized to 570 pixels high × 380 pixels wide for onscreen presentation. Experiments were run using a 21.5-inch iMac with i5 processor. Stimulus presentation and data collection were controlled by PsychoPy2 (Peirce et al., 2019).

To compare effects of race and familiarity directly, we constructed an Old/New recognition test in which the within-subjects factors of *Race* (*own*, *other*) and *Familiarity* (*familiar*, *unfamiliar*) were manipulated in a fully counterbalanced 2 × 2 factorial design. The experiment consisted of two main phases—a learning phase and a test phase. In the learning phase, participants viewed a series of 64 faces (16 black celebrities, 16 white celebrities, 16 black non-celebrities, 16 white non-celebrities) presented one at a time in a random order. In the subsequent test phase, participants viewed a longer series of 128 faces (32 black celebrities, 32 white celebrities, 32 black non-celebrities, 32 white non-celebrities) presented one at a time in a random order. Half of the faces in the test phase were *Old* faces that participants had seen in the learning phase. The other half were *New* faces that had not been presented before. Two complementary versions of the learning phase were counterbalanced across participants so that, when pooling over the whole experiment, each face appeared as an *Old* face and a *New* face an equal number of times.

We defined recognition performance as the proportion of correct responses in the memory test. To gauge metacognitive insight into the effects of each factor, we also recorded *self-estimates* and *peer estimates*. Self-estimates comprised three metrics. *Prospective* self-estimates were captured trial-by-trial in the learning phase. For each face, participants estimated the probability that they would recognise that face in the subsequent memory test (percentage response). *Concurrent* self-estimates were captured trial-by-trial in the test phase. For each face, participants rated their confidence that their own answer was correct (percentage response).

Retrospective self-estimates were captured at debrief. The 64 faces from the learning phase were presented in a single array, grouped according to the 2 × 2 factorial design (16 black celebrities, 16 white celebrities, 16 black non-celebrities, 16 white non-celebrities; cell positions counterbalanced). Participants were asked to reflect on the task as a whole, and to circle the two groups of faces they thought they remembered best. Choosing two of four options yields six possible combinations: 1–2, 1–3, 1–4, 2–3, 2–4, and 3–4.

Peer estimates were also captured trial-by-trial in the test phase. For each face, participants estimated (i) how many participants of the *same race* as the onscreen face would answer correctly, and (ii) how many participants of *different race* to the onscreen face would answer correctly (both estimates out of 30).

Procedure

The experiment began with the learning phase, in which participants made prospective self-estimates for each of the 64 faces using a percentile scale (figure 3.2d). Following the learning phase, participants completed a short filler task (number search) before proceeding to the test phase. For each of the 128 faces in the test phase, participants made four separate responses in a fixed order: (i) whether the face was *Old* or *New* (recognition response; figure 3.2d), (ii) their confidence that their own recognition response was correct (concurrent self-estimate; figure 3.2g), (iii) the number of *Same-race* participants they thought would give the correct recognition response (peer estimate; figure 3.2m), and (iv) the number of *Different-race* participants they thought would give the correct recognition response (peer estimate; figure 3.2m). After completing the test phase, participants viewed all 128 faces again, indicating whether or not they already knew each face before the experiment (participants did not have to know the person's name to know the person's face). These familiarity responses were used to define *familiar* and *unfamiliar* faces for each participant. The familiarity check was followed by the retrospective self-estimate (figure 3.2j). Finally, participants were asked to write down their own ethnic group (free response; figure 3.2j). These ethnicity responses were used to define own-race and other-race faces for each participant. The experimenter provided task instructions at the beginning of each task. No time limit was imposed for any of the tasks. Each display remained on screen until the participant's response, which immediately initiated the next trial. The entire test session took approximately 50 minutes to complete.

Recognition performance

We undertook three complementary analyses of recognition performance, based on signal detection measures, recognition accuracy (percentage of correct responses), and the percentage of participants who answered each item correctly. We also examined study time in the learning phase as a measure of encoding effort. Data from four participants (2 black and 2 white), whose accuracy fell 3 *SD* below the group mean, were excluded from analysis. In all four analyses, effects of familiarity were much stronger than effects of race.

Signal detection analysis. A 2 × 2 repeated-measures ANOVA of d' values revealed a significant main effect of *Familiarity*, with higher d' values for *familiar* faces ($M = 3.00$, $SE = .07$) than for *unfamiliar* faces ($M = 2.06$, $SE = .09$) [$F\ (1, 59) = 128.57$, $p < .001$, $\eta^2 = .33$]. The main effect of *Race* was also significant, with higher *d′* for *own*-race faces (*M* = 2.66, *SE* = .07) than for *other*-race faces (*M* = 2.40, *SE* = .07) [*F* (1, 59) = 26.55, *p* < .001, η^2 = .03].

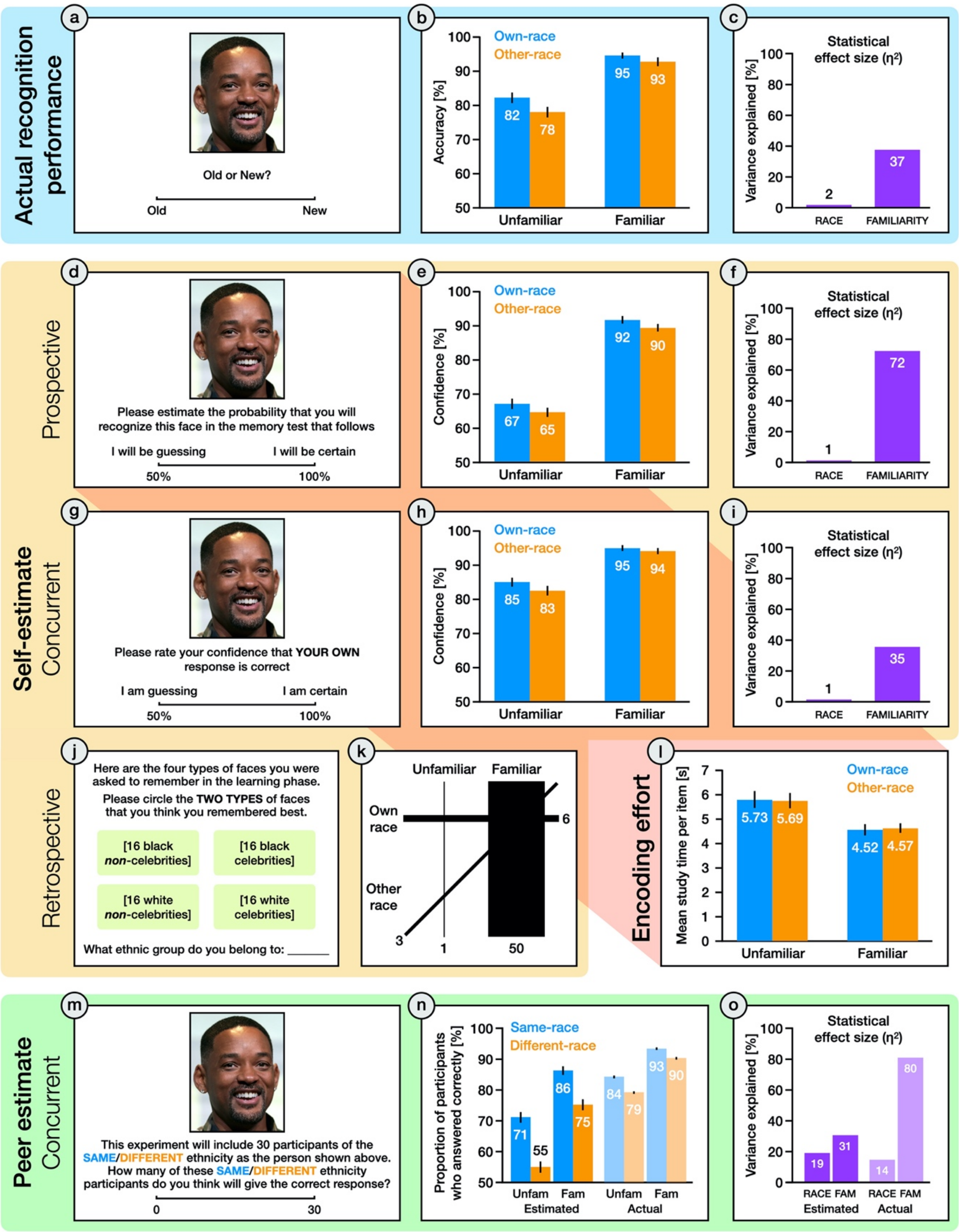
Table 3.2 **Signal detection analysis for Experiment 3.** Mean hit rates, false alarms, computed *d′* values, and Criterion *C* values in each condition. Standard errors are shown in parentheses next to the respective means.

	Own-race face	Other-race face
Hits		
Familiar face	.91(.01)	.89(.01)
Unfamiliar face	.68(.03)	.63(.03)
False alarms		
Familiar face	.05(.01)	.09(.01)
Unfamiliar face	.06(.01)	.09(.01)
d′		
Familiar face	3.13(.07)	2.87(.09)
Unfamiliar face	2.19(.09)	1.92(.09)
C		
Familiar face	.10(.03)	.09(.04)
Unfamiliar face	.51(.05)	.54(.06)

There was no significant interaction between these two factors [*F* (1, 59) = .01, *p* = .93, η^2 < .01]. A similar 2 × 2 ANOVA for Criterion *C* values also revealed a significant main effect of *Familiarity*, with a less strict criterion for *familiar* faces (*M* = .10, *SE* = .03) than for *unfamiliar* faces (*M* = .53, *SE* = .05) [*F* (1, 59) = 70.62, *p* < .001, η^2 = .26]. Participants were more likely to classify *unfamiliar* faces as New and *familiar* faces as Old. There was no significant main effect of *Race* [*F* (1, 59) =.09, *p* =.77, η^2 < .01], and no interaction between the two factors [*F* (1, 59) =.40, *p* =.53, η^2 < .01]. Descriptive statistics are shown in Table 3.2.

Accuracy. A 2 × 2 repeated-measures ANOVA with the factors of *Race* (*own*, *other*) and *Familiarity* (*familiar*, *unfamiliar*) also revealed a significant main effect of *Familiarity*, with higher accuracy for *familiar* faces (M = 93.92, SE = .71) than for *unfamiliar* faces (M = 80.35, SE = 1.24) [F (1, 59) = 131.04, p < .001, η^2 = .37]. The main effect of *Race* was also significant, with higher accuracy for *own*-race faces (M = 88.64, SE = .85) than for *other*-race faces (M = 85.62, SE = .95) [F (1, 59) = 17.12, p < .001, η^2 = .02]. There was no significant interaction between these the two factors [F (1, 59) = 3.26, p = .08, η^2 < .01] (figure 3.2b,c).

Proportion of participants who answered correctly. All proportions are expressed as percentages. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of *Familiarity*, with more participants giving correct responses to *familiar* faces (M = 91.18, SE = .12) than to *unfamiliar* faces (M = 81.70, SE = .17) [F (1, 59) = 4976.61, p < .001, η^2 = .80], and a significant main effect of *Race,* with more participants giving correct responses to *same*-race faces (M = 88.46, SE = .11) than to *different*-race faces (M = 84.43, SE = .15) [F (1, 59) = 8249.72, p < .001, η^2 = .14]. The interaction between *Familiarity* and *Race* was also significant [F (1, 59) = 80.10, p < .001, η^2 < .01]. Analysis of simple main effects showed that the effect of *Familiarity* was significant for both *same*-race faces [F (1, 59) = 2677.35, p < .001, η_p^2 = .98], and *different*-race faces [F (1, 59) = 5098.37, p < .001, η_p^2 = .99]. The simple main effect of *Race* was significant for both *familiar* faces [F (1, 59) = 854.87, p < .001, η_p^2 = .94] and *unfamiliar* faces [F (1, 59) = 5530.45, p < .001, η_p^2 = .99] (figure 3.2n,o).


Old or New?

Please estimate the probability that you will recognize this face in the memory test that follows

Please rate your confidence that **YOUR OWN** response is correct

Here are the four types of faces you were asked to remember in the learning phase.
Please circle the **TWO TYPES** of faces that you think you remembered best.

What ethnic group do you belong to: _______

This experiment will include 30 participants of the SAME/DIFFERENT ethnicity as the person shown above. How many of these SAME/DIFFERENT ethnicity participants do you think will give the correct response?

Figure 3.2 **Summary methods and results for Experiment 3.** Shaded regions show actual recognition performance (blue), self-estimates (yellow), and peer estimates (green). Example displays are shown on the left, with summary data on the right. Percentages are rounded to the nearest integer. Performance metrics show group means with standard error (SE). Statistical effect sizes show the proportion of the total variance attributable to each effect (η^2). (a) Actual recognition performance was assessed via Old/New judgements to *Old* and *New* faces [photo by Gage Skidmore CC BY-SA 3.0 (https://commons.wikimedia.org/wiki/File:Will_Smith_by_Gage_ Skidmore.jpg)]. (b) Percentage accuracy rates for each condition, and (c), statistical effect sizes for *Race* and *Familiarity* in the Old/New recognition memory test. (d) Prospective estimates of own recognition performance were assessed via confidence judgements for each face during the learning phase. (e) Percentage confidence ratings for each condition, and (f), statistical effect sizes for each factor in prospective estimates of own performance. (g) Concurrent estimates of own performance were assessed via confidence judgements for each face during the test phase. (h) Percentage confidence ratings for each condition, and (i), statistical effect sizes for each factor in concurrent estimates of own performance. (j) Retrospective estimates of own performance for each face type were captured after the recognition test. In the experimental display, all 64 faces from the learning phase were presented again, grouped as shown. Self-report ethnicity was used to define own-race and other-race faces for each participant in the analysis. (k) Combination plot showing the six possible ways to select two of the four face types in the retrospective self-estimate. Line thickness indicates frequency. (l) Mean study time per item for faces in each condition of the learning phase. (m) Concurrent estimates of peer performance for each face in the recognition test. Participants provided separate estimates for peers who shared the depicted race (*Same*) and peers who did not (*Different*). (n) Estimated and actual peer performance for each condition, and (o) statistical effect sizes for each factor, shown separately for estimated and actual peer performance.

Study time in learning phase. A 2 × 2 repeated-measures ANOVA showed a significant main effect of *Familiarity*, with shorter study times for *familiar* faces ($M = 4.51$, $SE = .20$) than for *unfamiliar* faces ($M = 5.71$, $SE = .32$) [F (1, 59) = 26.25, $p < .001$, $\eta^2 = .07$]. There was no significant main effect of *Race* [F (1, 59) < .01, p = .99, η^2 < .01], and no interaction between the two factors [F (1, 59) = .15, p = .70, η^2 < .01] (figure 3.2l).

Self-estimates

We conducted three separate analyses of self-estimates, based on *prospective* estimates (trial-by-trial confidence ratings in the learning phase), *concurrent* estimates (trial-by-trial confidence ratings in the test phase), and *retrospective* estimates (overall assessments at debrief). In all three analyses, effects of familiarity were much stronger than effects of race.

Prospective self-estimates. A 2 × 2 repeated-measures ANOVA with the factors of *Race* (*own*, *other*) and *Familiarity* (*familiar*, *unfamiliar*) revealed a significant main effect of *Familiarity*, with higher confidence for *familiar* faces (M = 90.71, SE = .79) than for *unfamiliar* faces (M = 65.97, SE = 1.10) [F (1, 59) = 555.49, p < .001, η^2 = .72]. The main effect of *Race* was also significant, with higher confidence for *own*-race faces (M = 79.54, SE = .87) than for *other*-race faces (M = 77.15, SE = .80) [F (1, 59) = 27.33, p < .001, η^2 = .01]. There was no significant interaction between the two factors [F (1, 59) =.06, p =.81, η^2 < .01] (figure 3.2e,f).

Concurrent self-estimates. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of *Familiarity*, with higher confidence for *familiar* faces (M = 94.58, SE = .63) than for *unfamiliar* faces (M = 83.76, SE = 1.05) [F (1, 59) = 139.14, p < .001, η^2 = .35]. There was also a significant main effect of *Race*, with higher confidence for *own*-race faces (M = 90.04, SE = .78) than for *other*-race faces (M = 88.30, SE = .79) [F (1, 59) = 10.20, p < .01, η^2 = .01].

Again, there was no significant interaction between the two factors [*F* (1, 59) = 2.03, p =.16, η^2 <.01] (figure 3.2h,i).

Retrospective self-estimates. Participants overwhelmingly chose *familiar, own-race* with *familiar, other-race,* indicating insight into the dominant effect of familiarity on their own performance. In contrast, the combination of *familiar, own-race* with *unfamiliar, own-race* was rarely chosen. A chi-square test confirmed that the frequencies for the six possible combinations were significantly different [χ^2 (5, N = 60) = 109.73, $p < .001$] (figure 3.2k).

Peer estimates

All proportions are expressed as percentages. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of *Familiarity*, with higher estimates for *familiar* faces (*M* = 80.65, *SE* = 1.23) than for *unfamiliar* faces (*M* = 62.97, *SE* = 1.31) [*F* (1, 59) = 253.24, $p < .001$, $\eta^2 = .30$]. There was also a significant main effect of *Race*, with higher estimates for *same-race* (*M* = 78.80, *SE* = 1.24) than for *different-race* (*M* = 64.82, *SE* = 1.45) [*F* (1, 59) = 95.63, $p < .001$, $\eta^2 = .19$]. The interaction between *Familiarity* and *Race* was also a significant [*F* (1, 59) = 47.70, $p < .001$, $\eta^2 = .01$]. Analysis of simple main effects showed that the effect of *Familiarity* was significant for both *same-race* faces [*F* (1, 59) = 186.70, $p < .001$, $\eta_p{}^2 = .76$], and *different-race* faces [*F* (1, 59) = 257.47, $p < .001$, $\eta_p{}^2 = .81$]. The simple main effect of *Race* was significant for both *familiar* faces [*F* (1, 59) = 69.34, $p < .001$, $\eta_p{}^2 = .54$], and *unfamiliar* faces [*F* (1, 59) = 107.00, $p < .001$, $\eta_p{}^2 = .65$] (figure 3.2n,o). In all our analyses of actual performance and self-estimates, *Race* had a small effect compared with *Familiarity*. For peer estimates, the two effects were of similar magnitude.

3.4.2 Experiment 4 (Different images at learning and test)

Participants

Sixty-two UK students (32 black, 30 white; 44 female, 18 male; mean age 22 years; age range 18–33 years) from the University of York took part in the experiment in exchange for a small payment or course credit. All participants provided written informed consent. None had participated in Experiment 3.

Stimuli and Design

The stimuli and apparatus were the same as in Experiment 3, except that we now presented a different image of each face in the learning phase and the test phase (figure 3.3d,a). The design and procedure were also the same as in Experiment 3, except that participants were now informed that different images of *Old* faces could be presented in the learning phase and the test phase.

Recognition performance

As with Experiment 3, we undertook three complementary analyses of recognition performance, using signal detection measures, recognition accuracy (percentage of correct responses), and the percentage of participants who answered each item correctly. We also examined study time in the learning phase as a measure of encoding effort. Data from two participants (both black), whose accuracy fell 3 *SD* below the group mean, were excluded from analysis. Once again, effects of familiarity were much stronger than effects of race in all four analyses.

Signal detection analysis. A 2 × 2 repeated-measures ANOVA of *d′* values revealed a significant main effect of *Familiarity*, with higher *d′* values for *familiar* faces ($M = 2.41$, $SE = .10$) than for *unfamiliar* faces ($M = .62$, $SE = .07$) [$F\ (1, 59) = 343.08$, $p < .001$, $\eta^2 = .60$]. The main effect of *Race* was also significant, with higher *d′* for *own*-race faces ($M = 1.69$, $SE = .07$) than for *other*-race faces ($M = 1.34$, $SE = .07$) [$F\ (1, 59) = 35.51$, $p < .001$, $\eta^2 = .02$].

Table 3.3 **Signal detection analysis for Experiment 4.** Mean hit rates, false alarms, computed *d′* values, and Criterion *C* values in each condition. Standard errors are shown in parentheses next to the respective means.

	Own-race face	Other-race face
Hits		
Familiar face	.87(.01)	.81(.02)
Unfamiliar face	.40(.03)	.38(.02)
False alarms		
Familiar face	.10(.01)	.16(.02)
Unfamiliar face	.18(.02)	.23(.02)
d′		
Familiar face	2.62(.10)	2.20(.12)
Unfamiliar face	.76(.08)	.48(.07)
C		
Familiar face	.08(.04)	.08(.05)
Unfamiliar face	.70(.07)	.59(.06)

There was no significant interaction between these two factors [$F\ (1, 59) = 1.29$, $p = .26$, $\eta^2 < .01$]. A similar 2 × 2 ANOVA for Criterion *C* values also revealed a significant main effect of

Familiarity, with a less strict criterion for *familiar* faces ($M = .08$, $SE = .04$) than for *unfamiliar* faces ($M = .65$, $SE = .06$) [F (1, 59) = 110.88, $p < .001$, $\eta^2 = .31$]. Participants were more likely to classify *unfamiliar* faces as New and *familiar* faces as Old. There was no significant main effect of *Race* [F (1, 59) = 2.15, $p = .15$, $\eta^2 < .01$], and no interaction between the two factors [F (1, 59) = 2.67, $p = .11$, $\eta^2 < .01$]. Descriptive statistics are shown in Table 3.3.

Accuracy. A 2 × 2 repeated-measures ANOVA with the factors of *Race* (*own*, *other*) and *Familiarity* (*familiar*, *unfamiliar*) also revealed a significant main effect of *Familiarity*, with higher accuracy for *familiar* faces ($M = 86.84$, $SE = 1.12$) than for *unfamiliar* faces ($M = 59.61$, $SE = .98$) [F (1, 59) = 489.16, $p < .001$, $\eta^2 = .68$]. The main effect of *Race* was also significant, with higher accuracy for *own*-race faces ($M = 75.33$, $SE = .93$) than for *other*-race faces ($M = 71.12$, $SE = .95$) [F (1, 59) = 29.28, $p < .001$, $\eta^2 = .02$]. There was no significant interaction between these the two factors [F (1, 59) = .54, $p = .46$, $\eta^2 < .01$] (figure 3.3b,c).

Proportion of participants who answered correctly. All proportions are expressed as percentages. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of *Familiarity*, with more participants giving correct responses to *familiar* faces ($M = 83.46$, $SE = .31$) than to *unfamiliar* faces ($M = 62.21$, $SE = .30$) [F (1, 59) = 3737.02, $p < .001$, $\eta^2 = .88$], and a significant main effect of *Race,* with more participants giving correct responses to *same*-race faces ($M = 75.87$, $SE = .25$) than to *different*-race faces ($M = 69.80$, $SE = .26$) [F (1, 59) = 22101.59, $p < .001$, $\eta^2 = .08$]. The interaction between *Familiarity* and *Race* was also significant [F (1, 59) = 96.61, $p < .001$, $\eta^2 < .01$]. Analysis of simple main effects showed that the effect of *Familiarity* was significant for both *same*-race faces [F (1, 59) = 4216.74, $p < .001$, $\eta_p^2 = .99$], and *different*-race faces [F (1, 59) = 2783.93, $p < .001$, $\eta_p^2 = .98$]. The simple main effect of *Race* was significant for both *familiar* faces [F (1, 59) = 2952.43, $p < .001$, $\eta_p^2 = .98$] and *unfamiliar* faces [F (1, 59) = 2811.24, $p < .001$, $\eta_p^2 = .98$] (figure 3.3n,o).

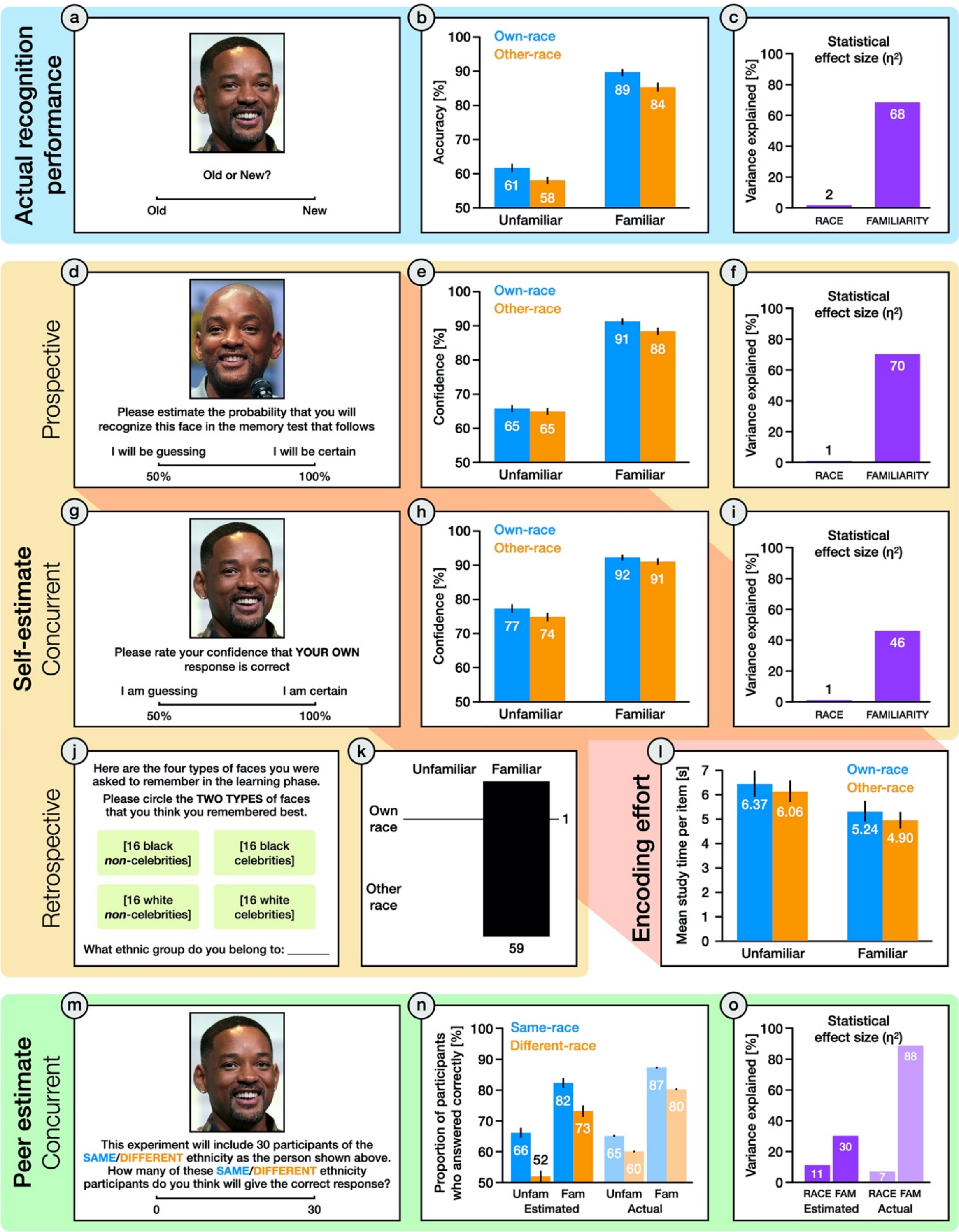


Figure 3.3 **Summary methods and results for Experiment 4.** (a) Actual recognition performance was assessed via Old/New judgements to *Old* and *New* faces. (b) Percentage

accuracy rates for each condition, and (c) statistical effect sizes for *Race* and *Familiarity* in the Old/New recognition memory test. (d) Prospective estimates of own recognition performance were assessed via confidence judgements for each face during the learning phase [photo by Gage Skidmore CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:Will_Smith_by_Gage_Skidmore_2.jpg]. (e) Percentage confidence ratings for each condition, and (f) statistical effect sizes for each factor in prospective estimates of own performance. (g) Concurrent estimates of own performance were assessed via confidence judgements for each face during the test phase. (h) Percentage confidence ratings for each condition, and (i) statistical effect sizes for each factor in concurrent estimates of own performance. (j) Retrospective estimates of own performance for each face type were captured after the recognition test. In the experimental display, all 64 faces from the learning phase were presented again, grouped as shown. Self-report ethnicity was used to define own-race and other-race faces for each participant in the analysis. (k) Combination plot showing the six possible ways to select two of the four face types in the retrospective self-estimate. Line thickness indicates frequency. (l) Mean study time per item for faces in each condition of the learning phase. (m) Concurrent estimates of peer performance for each face in the recognition test. Participants provided separate estimates for peers who shared the depicted race (*Same*) and peers who did not (*Different*). (n) Estimated and actual peer performance for each condition, and (o) statistical effect sizes for each factor, shown separately for estimated and actual peer performance.

Study time in learning phase. A 2 × 2 repeated-measures ANOVA showed a significant main effect of *Familiarity*, with shorter study times for *familiar* faces (M = 5.07, SE = .35) than for *unfamiliar* faces (*M* = 6.22, *SE* = .48) [*F* (1, 59) = 15.29, $p < .001$, $\eta^2 = .03$], and a significant main effect of *Race*, with shorter study times for *other*-race faces (*M* = 5.48, *SE* = .37) than for *own*-race faces (*M* = 5.81, *SE* = .44) [*F* (1, 59) = 4.58, $p < .05$, $\eta^2 < .01$]. There was no interaction between the two factors [*F* (1, 59) = .01, $p = .92$, $\eta^2 < .01$] (figure 3.3l).

Self-estimates

As for Experiment 3, we conducted three separate analyses of self-estimates, based on *prospective* estimates (trial-by-trial confidence ratings in the learning phase), *concurrent* estimates (trial-by-trial confidence ratings in the test phase), and *retrospective* estimates (overall assessments at debrief). In all three analyses, effects of familiarity were much stronger than effects of race.

Prospective self-estimates. A 2 × 2 repeated-measures ANOVA with the factors of *Race* (*own*, *other*) and *Familiarity* (*familiar*, *unfamiliar*) revealed a significant main effect of *Familiarity*, with higher confidence for *familiar* faces (*M* = 89.33, *SE* = .97) than for *unfamiliar* faces (*M* = 65.07, *SE* = .99) [*F* (1, 59) = 445.71, $p < .001$, $\eta^2 = .70$]. The main effect of *Race* was also significant, with higher confidence for *own*-race faces (*M* = 78.12, *SE* = .90) than for *other*-race faces (*M* = 76.28, *SE* = .86) [*F* (1, 59) = 15.48, $p < .001$, $\eta^2 < .01$]. The interaction between *Familiarity* and *Race* was also significant [*F* (1, 59) = 4.16, $p < .05$, $\eta^2 < .01$]. Analysis of simple main effects showed that the effect of *Familiarity* was significant for both *own*-race faces [*F* (1, 59) = 420.76, $p < .001$, $\eta_p^2 = .88$], and *other*-race faces [*F* (1, 59) = 340.31, $p < .001$, $\eta_p^2 = .85$].

The simple main effect of *Race* was significant for *familiar* faces [F (1, 59) = 16.62, $p < .001$, $\eta_p^2 = .22$], but not for *unfamiliar* faces [F (1, 59) = 1.73, $p = .19$, $\eta_p^2 = .03$] (figure 3.3e,f).

Concurrent self-estimates. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of *Familiarity*, with higher confidence for *familiar* faces (M = 91.15, SE = .84) than for *unfamiliar* faces (M = 75.68, SE = 1.21) [F (1, 59) = 204.58, $p < .001$, $\eta^2 = .46$]. There was also a significant main effect of *Race*, with higher confidence for *own*-race faces (M = 84.32, SE = .88) than for *other*-race faces (M = 82.51, SE = .96) [F (1, 59) = 17.16, $p < .001$, $\eta^2 = .01$]. There was no significant interaction between the two factors [F (1, 59) = 1.59, p =.21, $\eta^2 < .01$] (figure 3.3h,i).

Retrospective self-estimates. Participants overwhelmingly chose *familiar, own-race* with *familiar, other-race,* indicating insight into the dominant effect of familiarity on their own performance. In contrast, the combination of *familiar, own-race* with *unfamiliar, own-race* was rarely chosen. A chi-square test confirmed that the frequencies for the six possible combinations were significantly different [χ^2 (1, N = 60) = 56.07, $p < .001$] (figure 3.3k).

Peer estimates

All proportions are expressed as percentages. A 2 × 2 repeated-measures ANOVA revealed a significant main effect of *Familiarity*, with higher estimates for *familiar* faces (M = 77.76, SE = 1.63) than for *unfamiliar* faces (M = 58.69, SE = 1.58) [F (1, 59) = 227.38, $p < .001$, $\eta^2 = .30$]. There was also a significant main effect of *Race*, with higher estimates for *same-race* (M =

73.91, *SE* = 1.50) than for *different-race* (*M* = 62.53, *SE* = 1.68) [F (1, 59) = 88.99, $p < .001$, $\eta^2 = .11$]. The interaction between *Familiarity* and *Race* was also a significant [F (1, 59) = 34.57, $p < .001$, $\eta^2 = .01$]. Analysis of simple main effects showed that the effect of *Familiarity* was significant for both *same-race* faces [F (1, 59) = 225.91, $p < .001$, $\eta_p^2 = .79$], and *different-race* faces [F (1, 59) = 194.37, $p < .001$, $\eta_p^2 = .77$]. The simple main effect of *Race* was significant for both *familiar* faces [F (1, 59) = 62.70, $p < .001$, $\eta_p^2 = .52$], and *unfamiliar* faces [F (1, 59) = 92.84, $p < .001$, $\eta_p^2 = .61$] (figure 3.3n,o). In all our analyses of actual performance and self-estimates, *Race* had a small effect compared with *Familiarity*. For peer estimates, the two effects were of similar magnitude.

Chapter 4 Face-evoked thoughts

Reference:

Zhou, X., & Jenkins, R. (2020). Face-evoked thoughts. Manuscript submitted for publication.

Abstract

The thoughts that come to mind when viewing a face depend partly on the face and partly on the viewer. This basic interaction raises the question of how much common ground there is in face-evoked thoughts, and how this compares to viewers' expectations. Previous analyses have focused on early perceptual stages of face processing. Here we take a more expansive approach that encompasses later associative stages. In Experiment 5 (free association), participants exhibited strong egocentric bias, greatly overestimating the extent to which other people's thoughts resembled their own. In Experiment 6, we show that viewers' familiarity with a face can be decoded from their face-evoked thoughts. In Experiment 7 (person association), participants reported who came to mind when viewing a face—a task that emphasises connections rather than nodes in a social network. Here again, viewers' estimates of common ground exceeded actual common ground by a large margin. We assume that a face elicits much the same thoughts in other people as it does in us, but that is a mistake. In this respect, we are more isolated than we think.

4.1 Introduction

What comes to mind when you see a face? To some extent, it depends on the face—not only its physical appearance, but also the person whose face it is, and everything that goes along with that person. However, it also depends on the viewer. This is partly because different faces are known to different viewers. Some people know who Arnold Schwarzenegger is, and other people do not. But even among those who do, experiences and preferences can differ widely. A politician might think of Arnold first and foremost as the Governor of California, whereas a cinema-goer might think of him primarily as The Terminator (and both incarnations divide opinion for different reasons). This interplay between face and viewer raises the question of how much common ground there is in face-evoked thoughts. A face is a natural entry point to a social network, but if the same node can lead different viewers in different directions, it is not clear how much of the network is really shared.

Depth of processing becomes important here. For superficial aspects of face processing, such as deciding whether a person is female or male, observers' responses are highly consistent (Bruce et al., 1993; Burton, Bruce & Dench, 1993). Yet even the seemingly objective task of characterising face shape reveals wide discrepancies between viewers (Towler, White, & Kemp, 2014). Judgements of photographic likeness are similarly idiosyncratic (Hay, Young, & Ellis, 1991). There is not much consensus among viewers as to which photo best captures a person's appearance (White, Burton, & Kemp, 2016; Ritchie, Kramer, & Burton, 2018). All of these face processing tasks are superficial in the sense that observers need only consider the face as a physical surface. As cognition proceeds beyond physical cues to the inferences we draw from them, opportunities for divergence multiply. In an influential analysis of facial attractiveness, Hönekopp (2006) showed that private preferences are roughly as powerful as shared preferences in determining judgements. This finding came as something of a surprise, as it overturned the prevailing view at the time that agreement on such judgements among observers was high. But subsequent analyses have also concluded that private preferences, shaped by personal experience, are often the major determinant of attractiveness judgements (Germine et al., 2015; Hehman et al., 2017; Sutherland et al., 2020) and other trait inferences from faces—notably the cardinal dimensions of trustworthiness and dominance (Sutherland, Rhodes, Burton, & Young, 2020; Sutherland et al., 2020).

The preceding studies share some important features. All of them concern early stages of face processing. Converging evidence from ERP measures (e.g. Mouchetant-Rostaing et al., 2000) and saccadic reaction times (e.g. Ramon, Sokhn, & Calder, 2019) indicate that female and male faces can be differentiated within 150 ms of stimulus onset. Impressions of trustworthiness, dominance, and attractiveness based on facial appearance also emerge quickly—as early as 100 ms post-stimulus (Willis & Todorov, 2006; Olivola & Todorov, 2010). But the cascade of cognition that a face sets in motion lasts much longer. This is especially true for familiar faces. Interaction with semantic and emotional processes appears to peak around 400 ms and remains clear until at least 600 ms post-stimulus (Wiese et al., 2019). As stimulus associations are often idiosyncratic and often chain together (at least in the word domain; Shapiro, 1966, De Deyne et al., 2019), it is precisely in these later processes that one would most expect individual viewers to diverge. The resulting heterogeneity of responses can make analyses unwieldy. Perhaps for that reason—and undoubtedly because of applied interest in early stages of face processing (Brewer & Wells, 2011; Schultz, 2012; Phillips et al., 2018)—later stages of face processing have received less attention in cognitive research.

This brings us to a second commonality among previous studies. Responses are typically constrained to a small set of options—for example, whether a face is female or male, whether or not faces match on some dimension, or a numerical rating of a particular attribute. A participant's first thought when seeing a face might be, "She looks like my primary teacher!", but that reaction will not be captured as part of the data. There are some exceptions where researchers have gathered free associations to face images (e.g. Oosterhof & Todorov, 2008; Sutherland et al., 2018). However, in those cases, free associations were not the main interest. Instead, they were used to compute dimensions of variation for first impressions from faces. Experimental participants then rated faces on those dimensions using Likert scales.

Constraining responses in this way makes sense when the focus is a specific psychological mechanism. Our focus here is rather different, as our questions concern networks of social cognition. When it comes to face-evoked thoughts, little is known about the extent of overlap among viewers. But without direct insight into the minds of others, our sense of common ground can not reflect the actual extent of overlap. It can only reflect our *impression* of overlap, and that depends on certain metacognitive assumptions. What occurs to other people when they see a particular face? How does that compare with one's own experience? These questions have not been addressed experimentally, although other areas of psychology offer some important clues.

Across a wide range of situations, we are inclined to assume that others think as we do, sharing our tastes, preferences, and understanding (false consensus effects; Ross, Greene, & House, 1977; Krueger & Clement, 1994), and generally seeing the world from our own perspective. For example, we tend to overestimate the extent to which others know the things that we know (Hinds, 1999; Gilovich, Medvec, & Savitsky, 2000) and make the choices that we make (Ross, Greene, & House, 1977; Wolfson, 2000). These failures of metacognitive insight are examples of egocentric bias—the tendency to understand others from our own perspective (Krueger & Clement, 1994). Recently, egocentric biases have been demonstrated in face perception. In one example, participants in an identity matching task predicted that the faces they themselves knew would be easy for other people to match (Ritchie et al, 2015). Zhou & Jenkins (2020) showed that, in matching tasks for identity, emotional expression, and gaze direction, high performing participants attributed higher performance to other people, and low performing participants attributed lower performance to other people.

These findings demonstrate egocentric bias in early stages of face processing. Our scope here is deliberately more broad. Instead of focusing on early perceptual processes and fixed response options, we seek to capture whatever comes to mind when viewing the face. Importantly, this is not the same as establishing what the viewer knows about the seen person. At any moment, what comes to mind is only a subset of one's relevant knowledge. This is a critical distinction. It is what comes to mind, not what stays behind, that constitutes a train of thought. The current study addresses two related aspects of face-evoked thoughts—first, the degree of overlap among viewers, and second, how this overlap compares to viewers' expectations. Given the evidence of egocentric bias elsewhere in cognition, we predicted that viewers would overestimate the extent to which other people's thoughts resemble their own (a false consensus effect). We begin in Experiment 5 by asking viewers what comes to mind when they see a particular face. By focusing on the seen person, this question emphasises individual nodes in a social network. In

Experiment 6, we test whether face-evoked thoughts differ for familiar and unfamiliar faces. Finally, in Experiment 7, we ask *who* comes to mind when they see a particular face. By focusing on related people, this question emphasises connections between nodes.

4.2 Experiment 5

The purpose of this experiment was twofold. First, we sought to quantify overlap among observers in face-evoked thoughts. Second, we sought to compare the extent of this overlap to observers' expectations. To this end, we devised a face association task comprising cognitive and metacognitive components. Participants were asked to write down whatever thoughts came to mind when they viewed a series of faces (cognitive component). Analysing these responses allowed us to quantify actual overlap among observers. We then asked the same participants to estimate how other participants' responses would compare to their own (metacognitive component). Analysing these peer estimates allowed us to quantify expected overlap.

Our predictions concerned the numerosity, content, and order of face-evoked thoughts. At the cognitive level, we tested the following assumptions: (i) participants would differ in the number of thoughts they recorded, (ii) familiar faces would elicit more responses than unfamiliar faces, (iii) some responses to a particular face would be made by more than one participant, and (iv) salient associations would come to mind earlier than obscure associations. At the metacognitive level, we predicted the following signs of egocentric bias: (i) participants who produced a high number of responses would expect others to produce a high number of responses (and vice versa), (ii) participants would expect others to produce more responses for faces that they themselves knew, (iii) participants would overestimate the number of viewers who had the same thoughts that they themselves had, and (iv), this tendency to overestimate common ground would be strongest for thoughts that came to mind most readily.

4.2.1 Methods

Participants

Thirty UK students (25 female, 5 male; mean age: 19 years; age range 18–23 years) from the University of York took part in exchange for a small payment and task-related bonus. The experiments in this study were approved by the Ethics Committee at the University of York. All participants provided written informed consent.

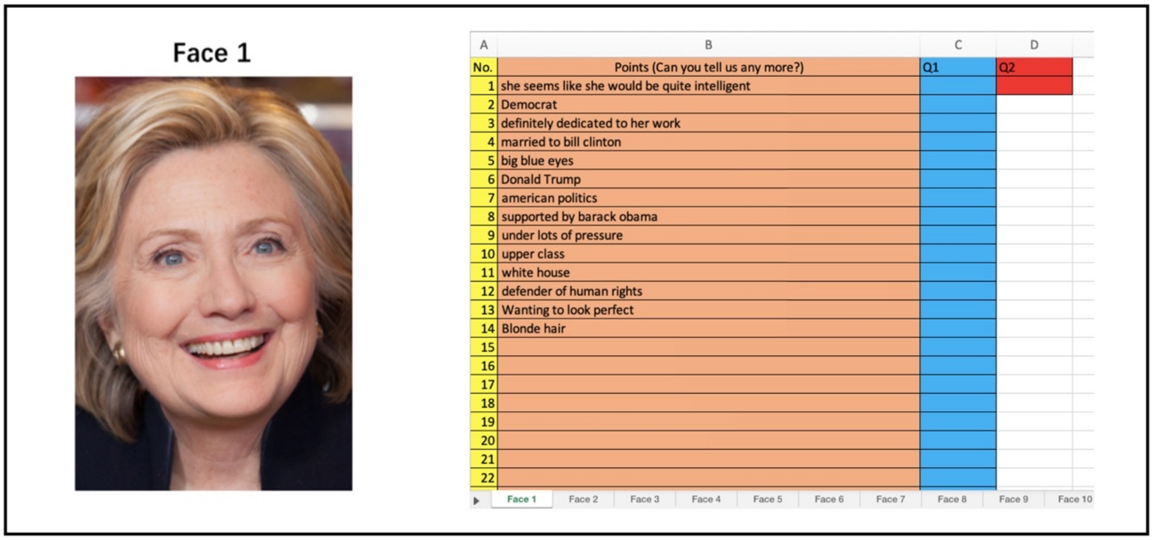
Stimuli and apparatus

Ambient images of 8 familiar celebrities (4 female; 4 male) and 8 unfamiliar celebrities (public figures from outside of the UK, 4 female; 4 male) were downloaded from the internet. The names of these celebrities are listed in Appendix B. Each photo captured the whole face with no occlusions from a roughly frontal aspect. In pilot testing, 64 UK students, who did not participate in the main experiment, indicated whether or not they were familiar with each face. An independent t-test confirmed that the familiar celebrities were known to more respondents (M = 89%) than the unfamiliar celebrities (M = 10%) [$t(14) = 14.99$, $p < .0001$].

All photos were cropped to 570 pixels high × 380 pixels and colour printed at 72 dpi onto A4 sheets, which were collated into booklets. Pagination was randomised, so that familiar and unfamiliar faces were intermixed. To counteract possible order effects, page order was reversed for half of the participants.

Design

Each participant completed three separate tasks—a face association test, a metacognitive review, and a face familiarity check. In the face association test, the participant's task was to capture whatever came to mind (their 'points') as they viewed each face. Participants transcribed their thoughts into a personalised Microsoft Excel workbook, using a separate worksheet for each face. Each worksheet was headed with the prompt, "Points (Can you tell us any more?)", followed by a series of enumerated rows to accommodate separate points of information (see Figure 4.1). This was a self-paced task, and participants were encouraged to be as exhaustive as possible in recording their thoughts. To motivate participants to generate as many points as possible, we introduced a cash incentive of 1p per point in addition to standard participant payments. For example, generating an average of 13 points for each of the 16 faces would result in an additional £2.08 payment.



No.	Points (Can you tell us any more?)	Q1	Q2
1	she seems like she would be quite intelligent		
2	Democrat		
3	definitely dedicated to her work		
4	married to bill clinton		
5	big blue eyes		
6	Donald Trump		
7	american politics		
8	supported by barack obama		
9	under lots of pressure		
10	upper class		
11	white house		
12	defender of human rights		
13	Wanting to look perfect		
14	Blonde hair		
15			
16			
17			
18			
19			
20			
21			
22			

Figure 4.1 Example stimulus (*left*) and response sheet (*right*) from the free association task in Experiment 5. For each face, participants wrote down whatever came to mind. Example response sheet shows genuine responses from different participants for illustrative purposes.

In the metacognitive review, participants revisited their own responses from the completed face association test. First, for each point they had made, participants were asked to estimate how many other participants (out of 30) would make the same point. Participants were instructed that the point didn't have to be expressed in exactly the same words, but should express the same idea. Second, for each face they had seen, participants estimated how many points other participants would generate on average. These data allowed us to compare participants' actual performance against their estimates of peers' performance on the same association task.

The face familiarity check was a computer-based task that was used to define familiar and unfamiliar faces for each participant. Participants were presented with the 16 stimulus faces one at a time in a random order. For each face, the participant's task was to indicate whether or not the face was familiar (Yes/No response). Each face remained on screen until the participant's keypress response, which initiated the next trial. Stimulus presentation and data collection were controlled by PsychoPy2 v1.82 (Peirce et al, 2019).

These three tasks gave rise to four types of data: (i) face associations—the thoughts that occurred to the participant upon seeing the face. These associations have numerosity, content, and sequential order; (ii) estimated overlap—for each point, the participant's estimate of how many other participants will make the same point; (iii) estimated numerosity—for each face, the participant's guess at how many points participants will generate on average; (iv) the participant's own prior familiarity with each face.

Procedure

All participants completed the face association test, then the metacognitive review, then the face familiarity check in the same fixed order. All three tasks were self-paced, and participants could take breaks at any time. The entire test session took approximately 45 minutes to complete.

4.2.2 Results and discussion

Numerosity

Numerosity refers to the number of points participants generated in the face association task. To test for egocentric bias in estimates of other people's performance, we divided participants into three equal-sized performance groups (*Low*, *Middle*, *High*) according to their overall numerosity scores. We then compared peer estimates for these groups, separately for Familiar and Unfamiliar faces. Summary data for each condition are shown in Figure 4.2A.

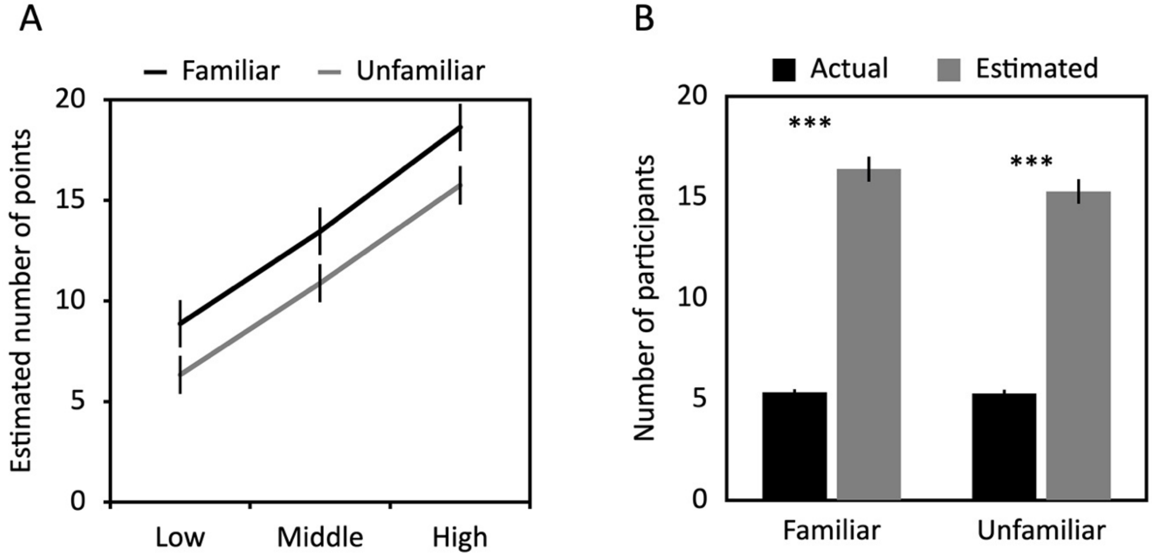


Figure 4.2 Egocentric bias and false consensus effects in the free association task in Experiment 5. (A) Peer estimates (condition means) from participants who themselves generated a low, middle, or high number of points, shown separately for familiar and unfamiliar faces. Peer estimates tracked participants' own performance. (B) Actual and estimated number of

participants (condition means) who made the same point to the same face, shown separately for familiar and unfamiliar faces. Estimated overlap exceeded actual overlap by a factor of 3. Asterisks indicate $p < .001$. Error bars show SE.

Participants' estimates were submitted to a 2 × 3 mixed ANOVA with the within-subjects factor of *Familiarity* (*Familiar*, *Unfamiliar*) and the between-subjects factor of *Group* (*Low*, *Middle*, *High*). This analysis revealed a main effect of *Familiarity*, with higher estimates for *Familiar* faces ($M = 13.66$, $SE = .68$) than for *Unfamiliar* faces ($M = 11.00$, $SE = .55$) overall [$F\ (1,27) = 30.82$, $p < .001$, $\eta p^2 = .53$]. There was also a main effect of *Group*, with estimates increasing from the *Low* group ($M = 7.61$, $SE = .99$) through the *Middle* group ($M = 12.18$, $SE = .99$) to the *High* group ($M = 17.20$, $SE = .99$) [$F\ (2,27) = 23.51$, $p < .001$, $\eta p^2 = .64$]. There was no significant interaction between these factors [$F\ (1,27) = .06$, $p = .95$, $\eta p^2 < .01$]. Participants expected others to generate more points for faces that they themselves knew, and fewer points for faces that they themselves did not know. Moreover, high scoring participants produced high peer estimates, and low scoring participants produced low peer estimates. Both of these findings point to egocentric bias in sizing up the face-evoked thoughts of other people.

Content

To quantify overlap among participants, we recruited two volunteer coders who categorised the face associations by content. We first grouped the data by face, pooling over participants, to create sixteen sets of associations (i.e. one set for each face). Each coder received all sixteen sets in a random order and independently worked through each set twice. On the first pass, coders classified each point as either physical (relating to appearance; e.g. "red hair") or non-physical (other information; e.g. "famous actor"). This classification gives us an indicator of abstraction from the visual domain. On the second pass, coders grouped together points that they judged to convey the same meaning (e.g. “famous actor”, “Hollywood star”), and assigned the size of the group to each point within the group. This procedure gives us the number of participants who made the same point to the same face, that is, the overlap among participants’ associations. Coders’ judgments were highly correlated [r (6297) = .71, p <.001]. Any point on which the coders differed was assigned the mean of the two group sizes.

Overlap refers to the number of participants who made the same point to the same face. Summary data for each condition are shown in Figure 4.2B. The overlap analysis was similar to the numerosity analysis. Overlap scores were submitted to a 2 × 2 repeated measures ANOVA with the within-subjects factors of *Familiarity* (*Familiar*, *Unfamiliar*) and *Measure* (*Estimated*, *Actual*). This analysis revealed a main effect of *Familiarity*, with higher scores for *Familiar* faces (M = 10.88, SE = .32) than for *Unfamiliar* faces (M = 10.30, SE = .32) overall [F (1, 29) = 5.96, p < .05, ηp^2 = .17]. There was also a significant main effect of *Measure*, with *Estimated* scores (M = 15.86, SE = .58) exceeding *Actual* scores (M = 5.32, SE = .16) overall [F (1, 29) = 310.12, p < .001, ηp^2 = .92]. These main effects were qualified by a significant interaction between *Familiarity* and *Measure* [F (1, 29) = 5.18, p < .05, ηp^2 = .15]. Simple main effects confirmed that *Estimated* overlap exceeded *Actual* overlap in both the *Familiar* condition [F (1, 29) = 295.14, p < .001, ηp^2 = .91] and the *Unfamiliar* condition [F (1, 29) = 244.18, p < .001, ηp^2 = .98]. The simple main effect of *Familiarity* was significant for *Estimated* score [F (1, 29) = 6.39, p < .05, ηp^2 = .18], but not for *Actual* score [F (1, 29) = .10, p = .76, ηp^2 < .01]. The main message from this analysis is that participants overestimated the degree of overlap between their own face associations and those of other people—a false consensus effect. Participants imagined that other viewers would have the same thoughts that they themselves had upon seeing a particular face. Such convergences did occur, but less often than participants expected.

Order

Associations come to mind sequentially. For this analysis, we assumed that order of occurrence is a proxy for association strength: strong associations come to mind early, weak associations come to mind later. The list format of participants' responses conserves this ordinal information. To ensure that our analysis remained representative of the participant group as a whole (N = 30), we excluded data beyond list position 21, where the total number of associations across participants fell below 30 (that is, below one association per participant). Figure 4.3A summarises the ordinal data. Spearman's correlations confirmed that early associations were more widely held than late associations, for *Familiar* faces [r (19) = -.94, p < .001] and *Unfamiliar* faces alike [r (19) = -.91, p < .001].

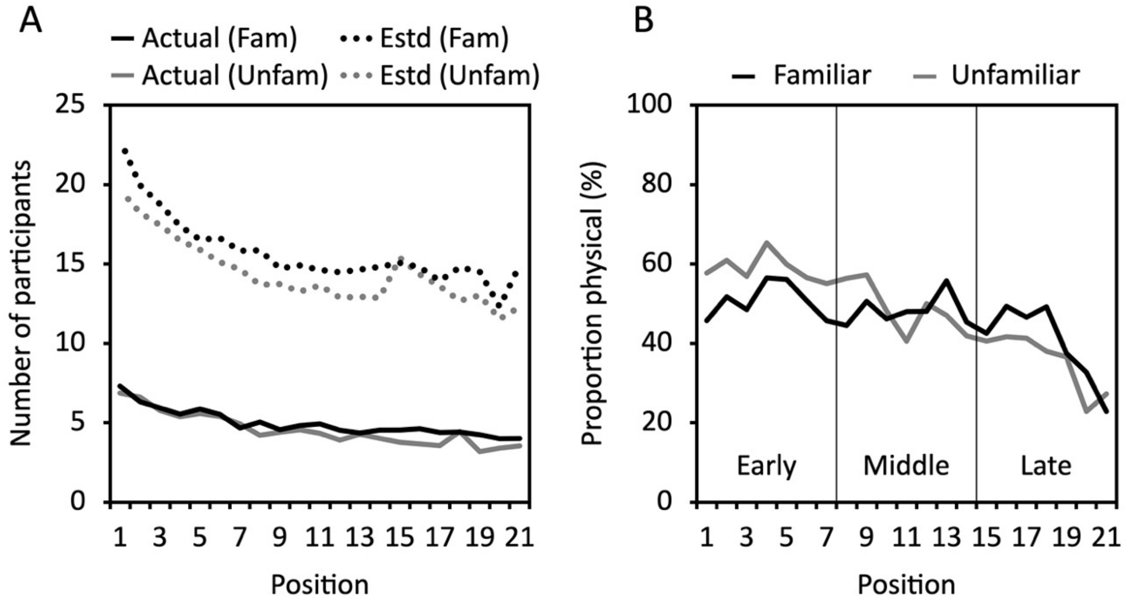


Figure 4.3 Ordinal analysis of consensus effects for the free association task in Experiment 5. (A) Actual and estimated (Estd) number of participants (condition means) who made the same point to the same face, plotted separately for familiar (Fam) and unfamiliar (Unfam) faces as a function of list position. Consensus was higher for earlier items than for later items. (B) Proportion of points that contained mainly physical information, plotted separately for familiar and unfamiliar faces as a function of list position.

Moreover, in keeping with egocentric bias, participants were more likely to attribute to others associations that came to mind early, and less likely to attribute to others associations that came to mind late (*Familiar* faces [r (19) = -.78, p < .001]; *Unfamiliar* faces [r (19) = -.84, p < .001]).

To test for qualitative differences in thoughts evoked by familiar versus unfamiliar faces, we next analysed participants' associations using the coders' classifications of content. Figure 4.3B shows the proportion of physical and non-physical associations as a function of list position, separately for *Familiar* and *Unfamiliar* faces.

To simplify the statistical analysis, we collapsed across list positions to form an *Order* factor with three levels—*Early* (positions 1–7), *Middle* (positions 8–14), and *Late* (positions 15–21). The proportion of points relating to physical information (facial appearance), as opposed to non-physical information (other associations), was analysed using a within-subjects ANOVA with the factors of *Familiarity* (*Familiar*, *Unfamiliar*) and *Order* (*Early*, *Middle*, *Late*). This analysis revealed no significant main effect of *Familiarity* [F (1, 6) = .77, p = .41, ηp^2 = .11]. However, there was a significant main effect of *Order,* with the highest proportion of physical points in *Early* responses (M = 54.80, SE = 1.42) followed by *Middle* responses (M = 48.55, SE = 1.43), and the lowest proportion in *Late* responses (M = 37.78, SE = 3.12) [F (1, 6) = 21.12, p < .001,

$\eta p^2 = .78$]. More importantly, there was a significant interaction between *Familiarity* and *Order* [$F\ (1, 6) = 12.66, p < .01, \eta p^2 = .68$]. Simple main effects revealed a significant effect of *Order* for both *Familiar* faces [$F\ (2, 24) = 7.16, p < .01, \eta p^2 = .37$] and *Unfamiliar* faces [$F\ (2, 24) = 31.86, p < .001, \eta p^2 = .73$]. The simple main effect of *Familiarity* was significant for *Early* responses [$F\ (1, 18) = 15.06, p < .01, \eta p^2 = .46$] and for *Late* responses (where the effect was reversed) [$F\ (1, 18) = 4.83, p < .01, \eta p^2 = .21$], but not for *Middle* responses [$F\ (1, 18) = .04, p = .89, \eta p^2 < .01$]. The content of face-evoked thoughts depends on the viewer's familiarity with the face. Physical information was more forthcoming for *Unfamiliar* faces than for *Familiar* faces. Conversely, non-physical information was more forthcoming for *Familiar* faces than for *Unfamiliar* faces.

One feature of participants' responses that we did not anticipate was spontaneous mention of other people's names. Evidently, a presented face often brought to mind another specific individual. This occurred occasionally for *Unfamiliar* faces (10.0% of responses), but significantly more often for *Familiar* faces (30.4% of responses) [$t(29) = 3.07, p < .01$]. This observation suggests that person associations could be among the most salient associations evoked by faces. We return to this finding in Experiment 7.

4.3 Experiment 6

Experiment 5 revealed a qualitative difference between thoughts evoked by familiar and unfamiliar faces. Thoughts concerning physical appearance came to mind more readily for unfamiliar faces than for familiar faces. Thoughts concerning non-physical attributes came to mind more readily for familiar faces than for unfamiliar faces. To follow up this finding, we investigated regularities between face familiarity and face associations in an independent test. We reasoned that if familiarity affects the content of face associations, it should be possible to judge (from associations alone) whether the person who made the associations was looking at a familiar face or an unfamiliar face. Moreover, if the balance of physical content is the basis for such judgements, then sorting associations by physical content should be equivalent to sorting them by familiarity, such that the two sorting tasks give rise to similar outcomes.

To test this possibility, we administered two separate categorisation tasks in which sorters reviewed participants' response sheets from Experiment 5. In one task, sorters judged whether each sheet contained mainly *Physical* information or mainly *Non-Physical* information (focus sort). In the other task, sorters judged whether the viewer was looking at a *Familiar* or *Unfamiliar* face (familiarity sort). We expected that response sheets that were categorised as *Physical* in the focus sort would be categorised as *Unfamiliar* in the familiarity sort. Conversely, response sheets that were categorised as *Non-Physical* in the focus sort should be categorised as *Familiar* in the familiarity sort.

4.3.1 Methods

Participants

Eighteen UK students (14 female, 4 male; mean age: 19 years; age range 18–25 years) from the University of York took part in exchange for course credit. None had participated in Experiment 1.

Stimuli and apparatus

The stimuli in this experiment were participants' response sheets from Experiment 5. Each of the 30 participants in Experiment 5 viewed 16 faces, resulting in a total of 480 response sheets. Screenshots of these response sheets were used as stimuli in the computer-based sorting tasks. Each image captured all of the associations that a single participant had made for a single face (see Figure 4.1). The same set of 480 images was used in the familiarity sorting task and the focus sorting task.

Design

To avoid carry-over effects, we randomly assigned the participants to one of two sorting tasks. Nine sorters categorised the response sheets by familiarity, and nine categorised them by focus. In the familiarity sort, their task was to judge, from the written associations on each sheet, whether the respondent was viewing a *Familiar* face or an *Unfamiliar* face. In the focus sort, their task was to judge whether the associations contained mainly *Physical* information or mainly *Non-Physical* information. These tasks allowed us to assign to each sheet two independent scores: (i) the number of times (out of nine) it was categorised as *Unfamiliar* (vs *Familiar*), and (ii) the number of times (out of nine) it was categorised as *Physical* (vs *Non-Physical*). Each participant saw all 480 response sheets in a random order.

Procedure

Participants received instructions for either the familiarity sort or the focus sort before completing the prescribed task. Each trial consisted of a single response sheet presented at the centre of the screen until response. Participants categorised each sheet via keypress (Q or P), which immediately triggered the next trial. The categorisation task was self-paced, and participants could take breaks at any time. The entire test session took approximately 30 minutes to complete.

4.3.2 Results and discussion

For each item in the categorisation tasks (i.e. each response sheet), we recorded the number of times it was categorised as *Unfamiliar* (familiarity score 0–9) and the number of times it was categorised as *Physical* (focus score 0–9). Figure 4.4 shows how many items received each combination of scores. As can be seen from the figure, familiarity scores and focus scores were strongly correlated [r (478) = .74, $p < .001$]. Specifically, *Unfamiliar* judgements cleaved with *Physical* judgements, and *Familiar* judgements cleaved with *Non-Physical* judgements. These regularities suggest that abstraction away from facial appearance is taken as evidence of familiarity.

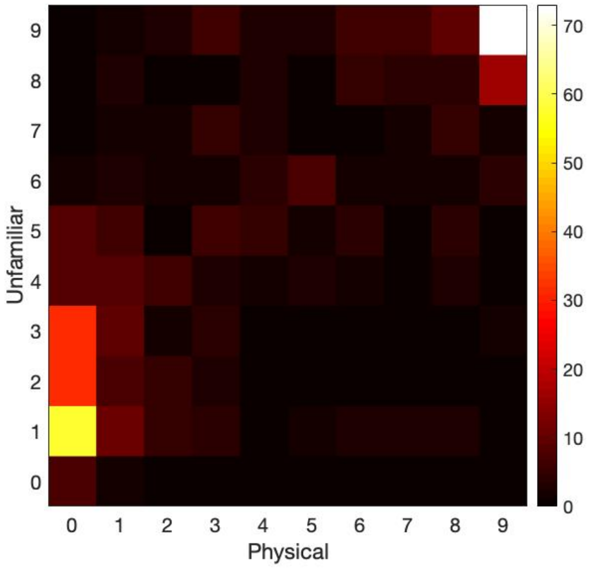


Figure 4.4 Analysis of face-evoked thoughts in Experiment 6. Responses that were deemed to be physical in content were deemed to refer to an unfamiliar face; responses that were deemed to be non-physical in content were deemed to refer to a familiar face. Colours indicate frequency.

To gauge the accuracy of these inferences, we next compared sorters' categorisations of familiarity (based on their reading the response sheets), to participants' actual familiarity with the faces concerned (familiarity checks in Experiment 5). The overall accuracy rate was 62%, significantly higher than chance performance of 50% [$z = 16.02$, $p < .001$]. With moderate accuracy, we can decode a person's familiarity with a face by reading what occurred to them when they saw it. In the final experiment, we focus on person associations evoked by faces, that is, connections between nodes in social networks.

4.4 Experiment 7

In Experiment 5, common ground among participants was smaller than participants expected. In some respects, the lack of common ground may not be surprising. After all, the face association task was entirely unconstrained, and we would expect some thoughts, such as episodic memories, or one's like or dislike of the depicted individual, to be idiosyncratic. However, not all face-evoked thoughts are idiosyncratic in this sense. One interesting aspect of Experiment 5 was participants' inclusion of personal names as associations with the seen face. Rather often, looking at the face led the viewer to think of someone else. This observation suggests that social associations are among the most salient thoughts that come to mind when viewing a face.

Semantic priming effects attest to the strength of such social associations. Viewers are typically faster to identify a known face when it is immediately preceded by a related person than when it is preceded by an unrelated person (Young et al., 1994). This phenomenon indicates that people who co-occur or share semantic information become closely associated in memory, such that activating the representation of one person activates the representation of related people (Burton et al., 1990; Wiese & Schweinberger, 2015). We suggest that this rings true when reflecting on everyday social encounters. Utterances such as, "Have you seen Mary?" or "How are the kids?" are common in conversation.

For present purposes, the key distinction is that social networks of co-occurrence and shared semantics are not idiosyncratic; they are objective features of the world. Although participants, by definition, do not share the same idiosyncratic associations, they do inhabit the same external world, albeit a particular corner of that world. This basic contrast raises the question of whether the pattern seen in Experiment 5 (overestimation of common ground) persists even when the association task is constrained to external connections between individuals.

If the observed pattern does persist, it would imply a more thoroughgoing egocentric bias: viewers wrongly assume that facts about social networks that occur to them also occur to others. Recognising that one's own perspective on the world is limited requires a leap of metacognitive insight. Recognising that another person's perspective will be different requires a further leap; and the process can fail at either stage. If the pattern is eliminated, this would imply a more limited egocentric bias: participants wrongly assume that face-evoked opinions that occur to them also occur to others, but they do not make the same mistake about face-evoked facts. Such a finding would suggest that a common frame of reference (the external world) allows better calibration of peer estimates.

To distinguish between these possibilities, we ran a modified version of the face association test in which associations were constrained to social relations, that is, connections between nodes in social networks. Instead of asking *what* comes to mind when viewing a face, we asked *who* comes to mind when viewing a face.

4.4.1 Methods

Participants

Thirty UK students (28 female, 2 male; mean age: 19 years; age range 18–21 years) from the University of York took part in exchange for a small payment and task-related bonus. None had participated in the preceding experiments.

Stimuli and apparatus

The stimuli and response booklets were the same as in Experiment 5, except that the instructions were modified to reflect the change in task.

Design

The study design was the same as in Experiment 5, except for the following changes. The face association test now called for associated people specifically, rather than any thoughts that came to mind. Accordingly, the header on each worksheet was modified to read, “Names (Can you tell us any more?)” (see Figure 4.5). Participants were instructed not to write the name of the person whose face was presented. Given that names can be difficult to recall, we accepted individuating semantic descriptions (e.g. “he played Harry Potter’s friend”) in cases where the name could not be retrieved or was never known. For simplicity, we include such entries as names in the rest of this paper. To discourage spurious responses, participants were also asked to supply a reason that the named individual was associated with the presented face.

A	B	C	D	E
No.	Names (Can you tell us any more?)	reason	Q1	Q2
1	Bill Clinton	Husband		
2	Donald Trump	That is who she is opposed with so is associated with.		
3	Theresa May	She is also a female politician who campaigned to be leader.		
4	Margaret Thatcher	Female politician		
5	My brother	She visited his University		
6	stephen colbert	interviewed her during her election campaign		
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Figure 4.5 Example stimulus (*left*) and response sheet (*right*) from the person association task in Experiment 7. For each face, participants wrote down whoever came to mind. Example response sheet shows genuine responses from different participants for illustrative purposes.

In the metacognitive review, participants were now asked to estimate, for each name, how many other participants (out of 30) would mention the same person, and for each face, how many names other participants would generate on average. These data allowed us to compare participants’ actual performance against their estimates of peers’ performance on the same association task.

Procedure

The procedure was the same as in Experiment 5. All participants completed the face association test, the metacognitive review, and the face familiarity check in that order.

4.4.2 Results and discussion

Numerosity

To test for egocentric bias, we again divided participants into three performance groups (*Low*, *Medium*, *High*) according to their overall numerosity scores. We then compared peer estimates for these groups, separately for *Familiar* and *Unfamiliar* faces. Summary data for each condition are shown in Figure 4.6A.

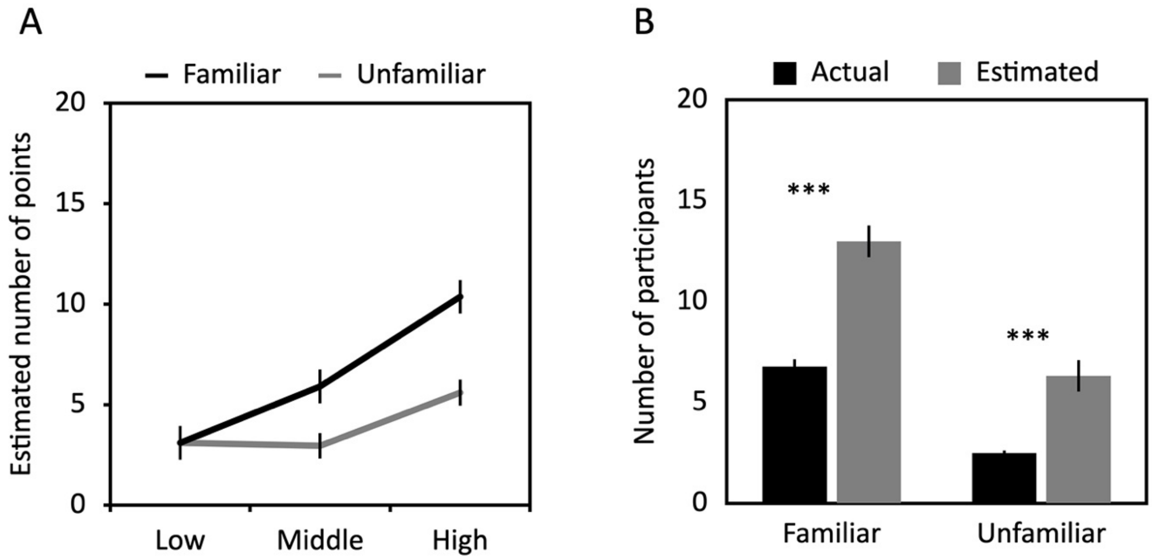


Figure 4.6 Egocentric bias and false consensus effects in the person association task in Experiment 7. (A) Peer estimates (condition means) from participants who themselves generated a low, middle, or high number of names, shown separately for familiar and unfamiliar faces. Peer estimates tracked participants' own performance. (B) Actual and estimated number of participants (condition means) who mentioned the same name to the same face, shown separately for familiar and unfamiliar faces. Estimated overlap exceeded actual overlap by at least a factor of 2. Asterisks indicate $p < .001$. Error bars show SE.

The analysis took the same form as before. Participants' estimates were submitted to a 2 × 3 mixed ANOVA with the within-subjects factors of *Familiarity* (*Familiar*, *Unfamiliar*) and the between-subjects factor of *Group* (*Low*, *Middle*, *High*). This analysis revealed a main effect of *Familiarity*, with higher estimates for *Familiar* faces ($M = 6.47$, $SE = .48$) than for *Unfamiliar* faces ($M= 3.89$, $SE = .37$) overall [$F\ (1,27) = 35.69$, $p < .001$, $\eta p^2 = .57$]. There was also a main effect of *Group*, with estimates increasing from the *Low* group ($M = 3.11$, $SE = .65$) through the *Middle* group ($M = 4.44$, $SE = .65$) to the *High* group ($M = 7.99$, $SE = .65$) [$F\ (2,27) = 15.28$, $p < .001$, $\eta p^2 = .53$]. These main effects were qualified by a significant *Familiarity* × *Group* interaction [$F\ (2,27) = 10.40$, $p < .001$, $\eta p^2 = .44$]. Simple main effects showed that *Familiar* estimates exceeded *Unfamiliar* estimates in the *Middle* group [$F\ (1,27) = 15.66$, $p < .001$, $\eta p^2 = .37$] and the *High* group [$F\ (1,27) = 40.83$, $p < .001$, $\eta p^2 = .60$], but not in the *Low* group [$F\ (1,27) < .01$, $p = 1.00$, $\eta p^2 < .001$]. The simple main effect of *Group* was significant for both *Familiar* faces [$F\ (2,27) = 19.15$, $p < .001$, $\eta p^2 = .59$] and *Unfamiliar* faces [$F\ (2,27) = 5.38$, $p < .05$, $\eta p^2 = .29$]. Overall, participants expected others to generate more names for faces that they themselves knew, and fewer names for faces that they themselves did not know. In addition, high scoring participants produced high peer estimates, and low scoring participants produced low peer estimates. The overall pattern is again indicative of egocentric bias, this time in estimating how many people will occur to other viewers when they see a particular face.

Content

In this analysis of content, overlap refers to the number of participants who mentioned the same name to the same face. Given that names are so constrained, matching responses were unambiguous. As such, the coding step in Experiment 5 was not necessary here. Summary data for each condition are shown in Figure 4.6B. As with the numerosity scores, overlap scores were submitted to a 2 × 2 repeated measures ANOVA with the within-subjects factors of *Familiarity* (*Familiar*, *Unfamiliar*) and *Measure* (*Estimated*, *Actual*). This analysis revealed a main effect of *Familiarity*, with higher scores for *Familiar* faces ($M = 9.89$, $SE = .51$) than for *Unfamiliar* faces ($M = 4.41$, $SE = .41$) overall [$F(1, 29) = 115.58$, $p < .001$, $\eta p^2 = .80$]. There was also a significant main effect of *Measure*, with *Estimated* scores ($M = 9.65$, $SE = .67$) exceeding *Actual* scores ($M = 4.64$, $SE = .21$) overall [$F(1, 29) = 62.75$, $p < .001$, $\eta p^2 = .68$]. These main effects were qualified by a significant interaction between *Familiarity* and *Measure* [$F(1, 29) = 10.65$, $p < .01$, $\eta p^2 = .27$].

Simple main effects confirmed that *Estimated* overlap exceeded *Actual* overlap in both the *Familiar* condition [$F(1, 29) = 80.73$, $p < .001$, $\eta p^2 = .74$] and the *Unfamiliar* condition [$F(1, 29) = 24.88$, $p < .001$, $\eta p^2 = .46$]. The simple main effect of *Familiarity* was significant for both *Estimated* score [$F(1, 29) = 66.81$, $p < .001$, $\eta p^2 = .70$] and *Actual* score [$F(1, 29) = 156.00$, $p < .001$, $\eta p^2 = .84$]. Once again, participants overestimated the degree of overlap between their own associations and those of other participants, this time, for social associations specifically. Participants imagined that other viewers would think about the same people that they themselves thought about upon seeing a particular face. In fact, the overlap was smaller than they expected.

Order

As with the free associations in Experiment 5, we analysed the order in which social associations were generated as a proxy for association strength. To ensure that our analysis remained representative of the participant group as a whole (N = 30), we excluded data beyond list position 5, where the total number of associations across participants fell below 30 (that is, below one per participant). Figure 4.7 summarises the resulting ordinal pattern.

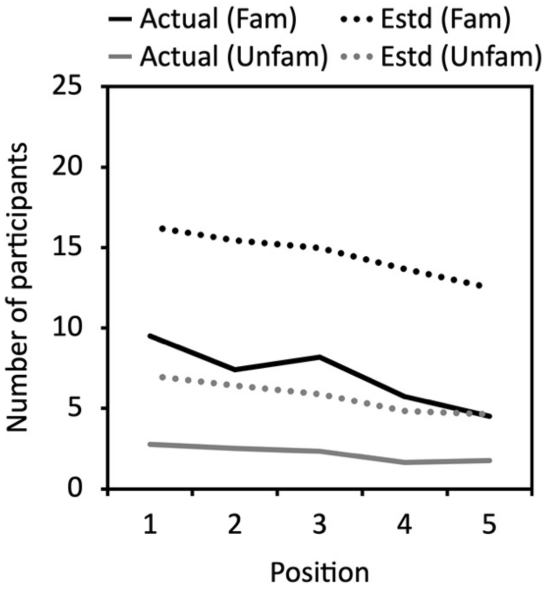


Figure 4.7 Ordinal analysis of consensus effects for the person association task in Experiment 7. Actual and estimated (Estd) number of participants (condition means) who mentioned the same name to the same face, plotted separately for familiar (Fam) and unfamiliar (Unfam) faces as a function of list position. Consensus was higher for earlier items than for later items.

Spearman's correlations confirmed that earlier associations were more widely held than later associations, for *Familiar* faces [r (3) = -.90, p < .05] and *Unfamiliar* faces alike [r (3) = -.90, p < .05]. As expected, participants were also more likely to attribute to others associations that came to mind early, and less likely to attribute to others associations that came to mind late (*Familiar* faces [r (3) = -.99, p < .001]; *Unfamiliar* faces [r (3) = -.99, p < .001]).

4.5 General Discussion

We set out to quantify overlap in face-evoked thoughts, and to compare the actual overlap with participants' expectations. This comparison revealed a consistent egocentric bias: across experiments, viewers overestimated the extent to which other people's thoughts resembled their own. In this respect, we are more isolated than we think.

These findings expand on previous work in a number of ways. First, they take an expansive view of face processing that runs from early perceptual stages through to late associative stages. In so doing, they shed new light on differences between familiar and unfamiliar face processing, contrasting the thoughts that they elicit in the viewer. Second, they encompass cognitive and metacognitive measures from the same participants. This approach illuminates the same processes from two different perspectives, and extends egocentric bias and false-consensus effects to a new area of social cognition.

Our cognitive measures conformed to our initial assumptions, providing a secure basis for comparison. For both free associations (Experiment 5) and person associations (Experiment 7), we observed large individual differences in the number of thoughts that participants recorded. Perhaps unsurprisingly, participants generated more responses to familiar faces than to unfamiliar faces. In addition, points that participants mentioned early were more likely to be mentioned by others. These quantitative differences were accompanied by qualitative differences in content. Physical information was especially forthcoming for unfamiliar faces, and non-physical information was especially forthcoming for familiar faces (Experiment 5). Naive observers were apparently sensitive to these regularities. In Experiment 6, sorters who categorised viewers' responses according to inferred familiarity with the face, and sorters who categorised the same responses according to their focus on physical information, produced similar solutions.

Our metacognitive measures revealed egocentric bias in both numerosity and content of face-evoked thoughts. For numerosity, participants who generated many responses expected other viewers to generate many responses, and vice versa. This egocentric bias tracked not only individual differences in participants' response rates, but also their familiarity with individual faces (Ritchie et al., 2015). For content, participants overestimated the number of viewers whose face-evoked thoughts matched their own—and by a large margin (cf. Bui, 2012). Peer estimates exceeded actual counts by a factor of 3 in Experiment 5, and by at least a factor of 2 in Experiment 7. These false-consensus effects gave rise to especially high estimates for thoughts that came to mind first.

All of these patterns were evident in a free association task that emphasised the seen person, corresponding to an individual node in a social network (Experiment 5). They were also evident in a person association task that emphasised related people, corresponding to connections between nodes in a social network (Experiment 7). Egocentric bias and false-consensus effects at both levels indicate that we overestimate common ground in face-evoked thoughts.

So far, we have discussed these findings in relative terms—estimated overlap exceeded actual overlap. However, it is also striking how small actual overlap was in absolute terms. Any given point was mentioned by only around 20% of participants (5 out of 30) on average, and even those associations that came to mind first did not command a majority. Moreover, 10% of free associations in Experiment 5 and 40% of person associations in Experiment 7 were unique, in that they were generated by only participant. At the time of writing, there is much speculation about social media and the segmentation of society into bubbles of like-minded people (Nikolov, Oliveira, Flammini, & Menczer, 2015; Nguyen, 2020). For this particular aspect of social cognition (face-evoked thoughts), false consensus appears so at odds with true consensus, it threatens to condemn each of us to a bubble of one. It may seem plausible, or even obvious, that communication allows us to escape this fate. After all, the whole purpose of communication is to improve insight into the minds of others (Fernbach & Sloman, 2017). However, cognitive biases are often deeply engrained and difficult to change (Kahneman, 2011). It is worth noting that our participants each brought to the experiment 20 years' experience in social cognition. Evidently, this everyday experience was not enough to quash egocentric bias in the social cognition tasks presented here. There is some evidence that egocentric bias diminishes with age (Yinon, Mayraz, & Fox, 1994; Hayashi & Nishikawa, 2019). Future experiments could test whether older adults become better attuned to those around them following their additional exposure.

The observed false consensus effects for face-evoked thoughts raise some interesting questions about own-face processing. People tend to be highly selective about photos of themselves (Hancock & Toma, 2009). Recent studies of photographic likeness and face identification suggest the operation of egocentric bias in selection of own-face photos. Specifically, skewed representations of self interfere with our ability to judge which photographs faithfully capture our own facial appearance (White, Burton, & Kemp, 2016; White, Sutherland, & Burton, 2017). Such findings concern primarily perceptual aspects of own-face processing. In light of the current findings, it would be interesting to test whether false consensus effects also emerge in associative aspects of own-face processing. If so, then people should expect their own face to evoke in others the same thoughts it evokes in them (Gilovich, Medvec, & Savitsky, 2000).

Chapter 5 General Discussion

5.1 Overview of findings

The research carried out in this thesis investigated cognition and metacognition in different aspects of face processing. Chapter 2 began by exploring whether the same Dunning–Kruger effects (DKE) and egocentric bias exist in familiar and unfamiliar face identity by modifying the procedure and analysis of the original Kruger and Dunning's (1999) and the classic same/different face matching paradigm (Experiment 1). Self-assessment measures (test score and percentile ranking) and peer assessment measures were taken for familiar and unfamiliar face matching tasks, separately. Dividing participants into four groups according to their actual performance, I found that low performers overestimated their own test score and percentile ranking in both familiar and unfamiliar face matching task, while high performers underestimated their own performance (except for test score in the familiar face matching task). As for peer assessment, top performers expected higher accuracy of other people than did low performers. Based on the consist findings of DKE and egocentric bias in face identity tasks, Experiment 2 explored whether the same metacognitive illusions also apply to the recognition of social signals from faces, and the stability across these tasks under the same paradigm format. The classic DKE and egocentric bias also emerged in emotional expression and gaze direction tasks (except for high performers' accurate self-assessment of test score). Participants' metacognitive performance was more stable than their cognitive performance across the three face recognition tasks. These findings demonstrate the existence of metacognitive illusions in different face perception tasks. The unskilled but unaware phenomenon should be given more attention. The stability of metacognition performance across different face tasks implies that there may be a common trait underlying estimates across tasks.

Despite there being no familiarity advantage in metacognition in Chapter 2, participants still performed better in familiar face identification than in unfamiliar face identification at the cognitive level. Given that familiarity has been widely showed to confer an advantage in face identification and recognition memory (Clutterbuck & Johnston, 2004; Noyes & Jenkins, 2017, 2019), the focus of the investigation in Chapter 3 next moved to the comparison of familiarity effect with another well-known effect in face memory, the other-race effect (ORE)— the finding that faces of one's own race are better remembered than faces of other race. Results of accuracy, signal detection analysis and the proportion of participants who answered correctly showed that the familiarity effect was at least three times larger than the other-race effect. Recogntion memory tasks based on identical images in the learning and test phase (Experiment 3) and different images of the same identity in each phase (Experiment 4) were combined with the prospective, concurrent and retrospective metacognitive measurements. The finding of the much stronger of familiarity effect was in line with participants' self-assessment. Interestingly, people overestimated the ORE when assessing other people's performance. The findings in this chapter demonstrate that the profile of familiarity effects and other-race effects in the face recognition literature is not in proportion to their importance as determinants of recognition accuracy. The observation that participants attributed more ORE to other people than to themselves is reminiscent of Blind Spot Bias (Pronin, 2007), the distinct condition of the egocentric bias as I reviewed in the General Introduction.

In Chapter 4, I next approached the investigation of egocentric bias from the face association angle. The experiments focused on three dimensions at both the cognitive and metacognitive level—numerosity (the number of points generated for each face), content (the number of participants who generated the same point) and order (the number of participants who generated the same point at each position). Clear egocentric bias emerged in each of these measures, whether the associations were free (Experiment 5) or were limited to names (Experiment 7). Those who generated many points or names expected that others would generate many, and those who generated fewer expected that others would generate few. Participants also overestimated the overlap between thoughts that occurred to them and thoughts that occurred to others. Moreover, they also expect this overlap to a larger extent for the early points then the late points they generated. A follow-up study (Experiment 6) about the content of face association showed that physical points were more related to unfamiliar faces and those abstracted beyond facial appearance were more related to familiar faces. Just by reading the person's associations, people could infer the familiarity of the face to the viewer.

In summary, according to the current findings, face is not a "special" domain in which deficits can be naturally intuited. The DKE and egocentric bias emerge in various aspects of face processing. These observations merit further attention from both basic and applied researchers.

5.2 Theoretical implications

Several theoretical implications follow from my research. In this section I go through each of them from the perspective of face perception and the perspective of metacognition separately.

5.2.1 Face perception

As captured in theoretical models (Bruce & Young, 1986, Burton et al., 1999), familiar face recognition has a performance advantage over unfamiliar face recognition. This has been known for many years cross a range of tasks (for review see Johnston & Edmonds, 2009, Young & Burton, 2017). Results of the current face matching study in Experiment 1 showed that identifying familiar faces (*M* = 95) was much easier than identifying unfamiliar faces (*M* = 82), which is consistent with the previous face matching studies (e.g., Bruce et al., 2001, Bruce et al., 1999, Burton et al., 1999, Megreya and Burton, 2006). Results of face memory experiments in this thesis showed that familiar faces (*M* = 94 in Experiment 3; *M* = 87 in Experiment 4) were remembered more accurately than unfamiliar faces (*M* = 80 in Experiment 3; *M* = 60 in Experiment 4) , which is consistent with previous face memory studies (e.g., Ellis et al., 1979, Klatzky and Forrest, 1984, Yarmey, 1971). This familiarity advantage was highlighted especially when put the relative magnitudes of race and familiarity effect in context. Based on the results in Chapter 3, I argue that previous research of face memory has put a lot of emphasis on the other-race effect. The observed three-fold dominance of familiarity effects over race effects does not diminish the importance of race, but does call for more attention to the importance of familiarity in fundamental research on face memory. Chapter 4 provides an initial theoretical account of the familiarity advantage in face-evoked thoughts. Previous research into face associations has rarely considered familiarity effects because it has focused on trait inferences in first impressions, which relate to unfamiliar faces only. The experiments in Chapter 4 examined effects of familiarity on anything and anybody that occurred to viewers when looking at a face. The results in Experiment 5 revealed no familiarity advantage in how many points people came up with (both $M = 13$) or how many people think about the same point (both $M = 5$). Interestingly, the familiarity advantage arose in the person association task in Experiment 7. Participants thought of more people when looking at a familiar face ($M = 5$) compared with an unfamiliar face ($M = 3$). Also, more participants came up with the same person when looking at the familiar face ($M = 8$) than at the unfamiliar face ($M = 2$). These findings call for future studies to understand the key differences between free association and person association in this context.

This present research is among the first to examine all the physical and non-physical information that comes to mind when looking at a face. Experiment 5 showed that people are more likely to firstly think about physical information to unfamiliar faces than to familiar faces, and that this trend reversed in the late points. Experiment 6 showed that if we look at the overall information people think about a face, unfamiliar faces were associated with mainly physical information while familiar faces were associated with mainly non-physical information. Furthermore, decoding whether a person is familiar with a face could rely on the person's overall associations with the face. This is an informative analysis, as the overall number of points was similar for familiar faces ($M = 14$) and unfamiliar faces ($M = 13$).

Along with new findings, the face association study also makes a contribution to the methodology of face perception. Unlike standard word association tasks, in which the participant is asked to give the first one or two words that come to mind after hearing the target word (e.g., Nissen & Henriksen, 2006; McNeill, 1966; Shapiro, 1966), there is no limit to the number of points participants could make in the face association task presented here. This makes for a much richer data set. The number of points and the order in which they were mentioned were both informative in Experiment 5 and Experiment 7. Previous research on first impressions (e.g., Penton-Voak et al., 2006) often asked participants to rate at a few given traits of a face. The current face association method has no restriction on the content people mention. In Sutherland et al.'s (2018) study, participants were allowed to write whatever came to mind when looking at a face, and they were also told it was a study about first impressions. Researchers analyzed their data by reducing the impressions that participants mentioned to cardinal dimensions, which were then used to collect ratings. Based on different aims, sorters in the current study (Experiment 5) were asked to sort participants' responses, but at two levels. One concerned general content (physical or non-physical). The other imposed no restriction, but required sorters to group the same meanings together, allowing us to count how many people mentioned the same point. One advantage of this approach is that the content of the general sorting level could be any categories of interest. Another methodological contribution of the current thesis is the same/different matching task for the emotional expression and gaze direction tasks. Matching has become a common paradigm for testing face identity performance. As I reviewed in Chapter 1, previous research on emotion expression and gaze direction often ask participants to choose the exact kind of emotion or gaze direction the face image shows. The same/different matching approach in the current work (Chapter 2) allows direct comparison of performance across identity, emotional expression and gaze direction in a task that has fixed demands, response options, and chance performance rate (50%). Importantly, the mean score across the three matching tasks were similar and all three revealed large individual differences (Experiment 2).

This method also provides another theoretical contribution, which is the relationship between face identity, emotional expression and gaze direction. It remains long-standing controversy in whether different aspects of face perception cleave together (e.g., Bruce & Young, 1986; Haxby et al. 2000) or interact with each other (e.g., Hadjikhani et al., 2008; Kaufmann & Schweinberger, 2004). Based on the same format of measurement in the thesis, the results provide more evidence supporting the view that invariant information (i.e., face identity) and changeable information (i.e., social signals such as emotional expression and gaze direction) are processed separately. Actual scores in the two changeable tasks (expression and gaze direction) were moderately correlated with each other, and neither was correlated with the invariant task (identity). From the perspective of overall performance, the results suggest that people who are good at face identity need not also be good at perception of changeable social signals from faces, and vice versa.

5.2.2 Metacognition

The experiments in this thesis are among the first to systematically examine metacognition in face processing. Together, they contribute to the literature in several ways. First and foremost, the current work is the first to explicitly link DKE and egocentric bias to face processing. The results in this thesis showed that face processing might not be such a special domain that peoplecan naturally realize their bias. The two metacognitive illusions did exist in both familiar and unfamiliar face matching tasks (Experiment 1), also in other changeable face perception tasks (Experiment 2). The egocentric bias in peer assessment also existed in estimates of the salience of one's own face-evoked thoughts (Chapter 4). The results of Chapter 3 also revealed a distinct form of egocentric bias, the bind spot bias, which is driven by an asymmetric consideration of evidence when evaluating oneself and others (Scopelliti et al., 2015). Specifically, people expected ORE to be stronger in other people than it actually was, but did not show this illusion in their own judgement. However, there are still some exceptions among top performers. Unlike in the classic DKE, top performers estimated their own test scores accurately rather than underestimating them in the familiar face identity, emotional expression and gaze direction matching tasks. Interestingly, performers in all four quartiles assessed their own performance at ceiling in those three tasks, with no cross over in the two changeable face perception tasks. Unlike the actual ceiling performance in the familiar face identity task (M = 96, see Figure 2.2A), the mean estimated score of participants in each quartile were all above the actual scores in the two social signal tasks (see Figure 2.5B,C). One possible reason why people were so overconfident in the social tasks is that they seldom receive direct and immediate negative feedback from others if they fail to observe others' emotion or interpret others' gaze direction, compared with their errors in face identity. Future studies could focus on effects of feedback effect on overconfidence in these tasks. Given some neuroscience evidence supported that the changeable and invariant aspects of face perception are represented in different brain regions (e.g., Bruce &Young, 1986; Haxby et al., 2000; Breen et al. 2000), it is also worth exploring whether this specialized module model also applies to metacognition performance.

The thesis also contributes to research exploring familiarity advantages in metacognition for several aspects of face perception. Interestingly, although familiar faces showed great advantage at the cognitive level of face identity, poor performers still overestimated their performance in a familiar face matching task (Experiment 1). The egocentric bias was also present in familiar face identity matching (Experiment 1), and estimates of the salience of face-evoked thoughts (Experiments 5 and 7). The strikingly different findings of familiarity advantage at the cognitive level and the metacognitive level remains an interesting direction for future study.

This thesis also makes methodological contributions to metacognition research. In Chapter 2, I tested perceived test score trial-by-trial instead of retrospectively (Kruger & Dunning, 1999). A number of metacognition experiments have used trial-by-trial measures in studies of emotional face recognition (Kelly & Metcalfe, 2011), other-race effect in face memory (Hourihan, Benjamin, & Liu, 2012; Rhodes, Sitzman, & Rowland, 2013) and movement memory (McIntosh et al., 2019). I used a 2-alternative force-choice (2AFC) question in the current trial-by-trial measure. Importantly, I did not directly ask participants whether they thought they got the answer right or wrong, such as asking whether they think they hit or missed the target in the movement and spatial memory task (McIntosh et al., 2019). Instead, I asked participants to judge whether they were sure or unsure with their answer, which was more in line with our actual situation when being asked to estimate our own performance. The DKE still emerged in this context (Chapter 2), which also to some extent verified the stability of DKE. However, the benefits of this measurement technique only apply to studies that aim to gain a specific estimated score to compare with the actual score. As introduced in Chapter 1 (see Table 1.2), confidence can also be a useful measure across different domains. In Chapter 3, a scale from 50% to 100% was used in both learning and testing phases of the face memory task, with a low point of 50% given that chance responding would result in 50% accuracy on the old/new 2AFC recognition test (see Rhodes et al., 2013, for a similar procedure in the field of metacognition in ORE). Besides the concurrent measure in Chapter 3, the prospective and retrospective measures were also used. Consistent results across multiple metacognitive measures provides stronger evidence in support of the conclusions. However, as I insisted in the General introduction, the corresponding measurement should be selected according to the research purpose. The prospective and concurrent measures are not suitable in the face association task. In Chapter 4, participants were encouraged to write as many face-evoked thoughts as they could, and were asked to make peer assessments only after completing this association phase. If they had known that each point they made required a peer estimate, that could potentially influence the number of points they wrote to some extent. They may also try to generate some more popular or common points and dismiss some special points. The retrospective measure I used in the current tasks avoids these concerns. The measurements on metacognition in face association still need to be further explored but the three dimensions used in the current work (numerosity, content and order) provide a useful precedent.

5.3 Practical implications

The practical implications of the cognition and metacognition findings are deeply interconnected for several application contexts. Here, I discuss each implication according to the contexts of social interaction, clinical treatments and security and forensic settings.

5.3.1 Social interaction

Chapter 2 revealed large individual differences in different aspects of face perception. High performers presumed that others also performed well, according to their own excellent ability. If they could understand that others may be unskilled but unaware, misunderstandings may be reduced accordingly. On the other hand, those who perform well in one face task do not necessarily perform well in other face tasks (Experiment 2). Once people know this fact, they may better understand how someone could recognize them rapidly even they only met once before yet could not read their emotions correctly. In the field of face memory, Chapter 3 showed that people over-attributed the ORE in other people. This misjudgement could also be somewhat eliminated if they realize the fact that the accuracy of other people's face memory is determined more strongly by knowing the person's face than by sharing the person's race. Another phenomenon of concern is that people tend to overestimate the overlap between face-evoked thoughts that are salient to them, and face-evoked thoughts that are salient to other people. This was the finding for both familiar and unfamiliar faces (Chapter 4). This finding provides evidence of false consensus effects (Ross et al., 1977), which refers to the tendency to assume that one's own opinions and beliefs are relatively widespread through the general population. It is perhaps understandable that people tend to seek approval from others and would like to make friends with those who share the same opinions and beliefs with themselves.

Nevertheless, in other research (Dean, 2007), a companion bias emerged in people's attitudes to those who did not share their choice. They assumed those who don't agree with them are defective in some way, which is detrimental to social interaction.

5.3.2 Clinical treatments

Chapter 2 showed a significant "unskilled but unaware" pattern in low performers in multiple face perception tasks including familiar and unfamiliar face identity, emotional expression recognition and gaze direction recognition. This is in line with the previous research in prosopagnosia showing that people with developmental prosopagnosia sometimes have little knowledge of their impaired facial recognition abilities (e.g., Fine, 2012). Similarly, people with autism spectrum disorders (ASD) may not realize their lack of skills in reading social signals from faces (e.g., Bishop & Seltzer, 2012). As I mentioned in the General Introduction chapter, knowledge of knowledge can be as important as knowledge itself. If people can not be aware when they have more trouble than others in face perception, it is unlikely that they will seek for help (Yardley et al., 2008). The severe impact of social interaction could also lead to psychological consequences, such as the increased level of anxiety (Diaz, 2008). Thus, the current work suggests that a lack of awareness of the disorder is the premise of the disorder, which should be emphasised in clinical treatments for face perception deficits.

5.3.3 Security and forensic settings

Given that face perception plays an important role in security and forensic settings, this work also has some practical implications in these areas. Eyewitness testimony is known to be unreliable, especially when the witness and suspect are of different races. In fact, eyewitness misidentification is the leading reason that innocent people are convicted (Scheck et al., 2000, Wells et al., 2006). In this context, the other-race effect plays a role that cannot be neglected. According to the meta-analysis of Meissner and Brigham (2001), eyewitnesses were 1.56 times more likely to misidentify in the other-race condition than in the same-race condition. Increased misidentification caused by race was also evident in the current thesis, both in an easy task with the same image in the learning and testing phases (Experiment 4) and in a hard task with different images in each phase (Experiment 5). However, the current work also revealed that when comparing the relative magnitude of ORE and familiarity effects, familiarity showed at least a three-fold dominance over race (Chapter 3). In other words, when witnessing a crime committed by someone with whom one is familiar (e.g., neighbors, colleagues and friends), an eyewitness identification could be somewhat reliable, even if the suspect and the eyewitness are of different races. Conversely, if the suspect is unfamiliar to the witness, identification is less likely to be reliable, irrespective of race. Until now, the relative impact of these two factors on face recognition performance has not been clear. People tend to over-attribute ORE, according to the peer assessment results in Chapter 3. Certainly, much more reality factors should be considered in the real-world eyewitness identifications, such as the decision time (Grabman et al., 2019), high stress (Deffenbacher et al. 2004) and individual differences in face identification ability (Bindemann et al., 2012). Therefore, this remains an interesting avenue for further field studies and the analysis of real cases if it is also possible to put the relative magnitude of the two factors in context.

5.4 Future directions

This thesis extends research on metacognition into face processing and has discussed the relationship between the face processing level and the metacognition level in several aspects. However, there is still a lot more to learn in this field based on the evidence from this thesis. Besides what I have mentioned in this chapter, the first interesting question is what leads to the consistency of metacognition performance across different face matching tasks, despite the inconsistency of the performance at the cognition level. One possible direction is to explore whether this is a phenomenon that is specific to the face domain, or whether it is a domain general pattern. That is, to compare performance in face and non-face domains using a within-subject design. My pilot work showed some potential evidence supporting the domain-general possibility. Participants were asked to complete both an unfamiliar face matching task (similar to the task in Chapter 2) and a flag and country matching task that used the same 2AFC paradigm. The results showed no correlation in actual performance across the two tasks, but a moderate correlation in their estimated performance [$r(44) = .39$, $p < .01$]. However, previous metacognitive research on intelligence has found that metacognitive skills are domain specific (Veenman & Spaans, 2005). Additional work concerning more detailed comparisons are needed to clarify the domain specificity or domain generality of metacognitive estimates. Another interesting perspective is whether there is any identifiable traits that underlie consistent performance estimates across different tasks. The effect of traits on self-assessments has been reported in other non-face domains (e.g., Ames & Kammrath, 2004; Soh & Jacobs, 2013). It is possible that underlying traits also play an important role in self-assessment across different aspects of face perception. Some related personality measures could be combined with face tasks in the future studies to see whether the same trait would predict metacognitive performance across several aspects in face perception. Furthermore, if consistency of metacognitive performance emerges across face tasks and non-face tasks, it would be interesting to test whether the same traits also predict metacognitive performance across face perception and other cognitive domains.

Another unanswered question that emerges from the current work is how to improve metacognitive performance. How to avoid adverse consequences arising from the DKE and egocentric bias is a topic worth discussing. Are metacognitive biases in face processing amenable to training? Top performers underestimate their own performance relative to that of others. This misperception could potentially be corrected by simply showing them the responses of other people. The same intervention has no effect on poor performers, whose error is to overestimate their own skill (Kruger & Dunning, 1999). In pilot work, I found a similar pattern for both low and high performers in a matching task for unfamiliar faces. For familiar face recognition, I expect that training may improve estimates of a particular person's performance (for example, a person who is seen to recognize footballers but not musicians). However, such training effects should not generalize because different people know different faces. Perhaps more importantly, how can we improve the metacognitive performance of low performers who are unskilled but unaware? Different studies have examined improvement of metacognition, but so far, there is no general consensus. Some researchers put the emphasis on the importance of the cognition level. For example, in a logical reasoning task, Kruger and Dunning (1999) showed that training low performers' cognition skills to a competitive level (to make them high performers) could improve their metacognition skills. However, no related evidence has been found in the improvement of metacognition in face identity research. For example, Alenezi and

Bindemann (2013) found no significant effect of trial-by-trial feedback. Conversely, by giving the feedback while the face remained onscreen, White et al. (2014) found that accuracy increased by 10%. As for professional training courses, Towler et al. (2019) found that professional facial image comparison training courses did not improve identification accuracy. Towler, White and Kemp (2020) found very little evidence that training the core face recognition system (e.g., the holistic perception processing) could improve face recognition or face matching accuracy based on evidence from training for facial image comparison practitioners, prosopagnosia patients and the general population. However, they encouraged researchers to further explore training methods from the feature-based route, that is, to promote feature-based strategies for extracting identity information from faces. These strategies are related to the cognition of cognition which is at the meta-level. Some researchers in non-face fields also stressed the importance of improving the metacognition level. They found that errors of overconfidence occurred at the metacognition level, instead of the memory level in a metamemory study (Cavanaugh & Perlmutter, 1982). In addition, Butler, Karpicke & Roediger (2008) suggested that giving feedback could help correct metacognitive error based on two experiments of testing general knowledge facts. Further research is necessary to explore whether it can also apply to the domain of face perception. In addition to investigating solutions for debiasing self-assessments from both cognitive and metacognitive perspectives, ways to improve peer assessments also remain unknown. Krueger and Clement (1994) noted that egocentric bias was ineradicable even after standard debiasing strategies such as feedback and education. Since the egocentric bias refers to the fact that people would estimate other people's performance according to their own, if people also made false judgement of their own performance, how would they estimate others' performance correctly? Thus, whether peer assessment could be debiased by improving self-assessment could be an interesting direction for future research.

Another interesting question based on the findings from the current work is whether metacognitive judgement influences subsequent behavior. For example, if those who make higher self-assessments pay less attention to face perception or have less desire to improve relevant strategies in their daily life, compared with those with low confidence in their own face perception ability. Once low performers are aware of their incompetence in face perception, will they ask for help or spend more time on learning related strategies? When people feel low confidence in assessing other people's social signals, will they ask for interpretation directly or remain silent? According to Nelson and Narens (1990), regulating some aspect of cognitive activity (metacognitive control) follows from assessing its current state (metacognitive monitor; see Table 1.1). Metacognitive control has been widely explored in learning and education (e.g., Azevedo, 2005; Anderson et al., 2006; Roebers et al., 2009). There is now an opportunity for researchers to address the significant knowledge gap in the domain of face processing, not just to reveal the current assessment accuracy, but to improve our understanding of the way it can influence subsequent behavior.

The current work also provide some points for research on face recognition algorithms. Studies of how humans perceive faces have been used to help design machine-based face recognition systems, which have already been deployed in many practical systems, e.g., at ports of entry at international airports in Australia and Portugal (see the review of face recognition by computers and humans at Chellappa, Sinha & Philips, 2010). With the rapid development of face recognition algorithms, more and more evidence shows that computers surpass humans on face matching in several conditions (e.g., Phillips et al. 2007; O'Toole et al. 2007; Tang & Wang, 2004), and fusing humans and algorithms can lead to near-perfect accuracy (O'Toole et al. 2007). According to Towler, Kemp and White (2017), unfamiliar face matching in applied settings, such as photo ID issuance and forensic investigation, are made by chains of humans and computers together. For example, the investigator will firstly submit facial imagery of a suspect. Face matching algorithms then search the database and return highly similar identities to a facial reviewer, who then sends several potential matches to the investigator. Importantly, the step where a human checks the output of the algorithms' database search can also lead to identification errors, even up to a 50% error rate, according to the results of the real-world task from White et al. (2015). On the other hand, face recognition algorithms also make errors. Given that humans show an ORE for face recognition, recent research also revealed a race bias in the face recognition algorithms (see Cavazos et al., 2020, for the review). For example, Phillips et al. (2011) found that algorithms developed in Western countries perform more accurately for Caucasian faces, whereas algorithms developed in East Asia perform more accurately for East Asian faces. The current work puts forward new directions for these phenomena from the perspective of metacognition, specifically, whether there is some bias when facial reviewers check the output from algorithms. For example, Chapter 3 revealed that people over-attribute the ORE in the performance of their peers. This raises the question of whether facial reviewers also overestimate race bias when assessing the performance of algorithms on faces of a minority race.

In sum, the current thesis has introduced metacognition to the field of face processing. It has not only enriched our psychological understanding of metacognition in different aspects of face processing, but also brought new insights into social interaction, forensic and other application settings. For the future, it would be good to see more developments start with these psychological issues and build technologies around them.

Appendices

Appendix A. List of Celebrities in Chapter 3

Black Celebrities

Counterbalancing Version A. Alicia Keys, Beyoncé, Cardi B, Jada Pinkett Smith, Mel B, Oprah Winfrey, Serena Williams, Zendaya, Barack Obama, Drake, Dwayne Johnson, Idris Elba, Kendrick Lamar, Morgan Freeman, The Weeknd, will.i.am.

Counterbalancing Version B. Ariana Grande, Jennifer Hudson, Kelly Rowland, Mariah Carey, Michelle Obama, Nicki Minaj, Rihanna, Whitney Houston, Eddie Murphy, Jay-Z, Kanye West, Kevin Hart, Michael B. Jordan, Samuel L. Jackson, Travis Scott, Will Smith.

White Celebrities

Counterbalancing Version A. Angelina Jolie, Ellen DeGeneres, Emma Watson, Hillary Clinton, Julia Roberts, Katy Perry, Kristen Stewart, Miley Cyrus, Benedict Cumberbatch, Chris Hemsworth, David Beckham, Harry Styles, Justin Bieber, Leonardo DiCaprio, Rowan Atkinson, Rupert Grint.

Counterbalancing Version B. Theresa May, Emma Stone, Scarlett Johansson, Selena Gomez, Anne Hathaway, Adele, Lady Gaga, Taylor Swift, Brad Pitt, Daniel Radcliffe, Donald Trump, Ed Sheeran, Gary Barlow, Martin Freeman, Robert Downey Jr., Ryan Reynolds.

Appendix B. List of Celebrities in Chapter 4

Familiar celebrities. Andrew Lincoln, Avril Lavigne, David Beckham, Hillary Clinton, Mark Zuckerberg, Rupert Grint, Taylor Swift, Theresa May.

Unfamiliar celebrities. Alexander Becht, Daniele Pecci, Helen Dalley, Mathias Lauridsen, Mélanie Laurent, Minka Kelly, Stuart Bellamy, Yana Marinova.

References

Achtziger, A., Martiny, S. E., Oettingen, G., & Gollwitzer, P. M. (2012). Meta-cognitive processes in the self-regulation of goal pursuit. In P. B. & K. DeMarree. (Eds.), *Social Meta-cognition. frontier of social psychology series* (pp. 121–139). New York, NY: Psychology Press

Alenezi, H. M., & Bindemann, M. (2013). The Effect of Feedback on Face-Matching Accuracy. In *Applied Cognitive Psychology* (Vol. 27, Issue 6, pp. 735–753).

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804–825.

Ames, D. R., & Kammrath, L. K. (2004). Mind-Reading and Metacognition: Narcissism, not Actual Competence, Predicts Self-Estimated Ability. *Journal of Nonverbal Behavior*, *28*(3), 187–209.

Anderson, M. L., Oates, T., Chong, W., & Perlis, D. (2006). The metacognitive loop I: Enhancing reinforcement learning with metacognitive monitoring and control for improved perturbation tolerance. *Journal of Experimental & Theoretical Artificial Intelligence: JETAI*, *18*(3), 387–411.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, *23*(3), 905–913.

Anson, I. G. (2018). Partisanship, political knowledge, and the Dunning-Kruger effect. *Political Psychology*, *39*(5), 1173–1192.

Arizpe, J. M., Saad, E., Douglas, A. O., Germine, L., Wilmer, J. B., & DeGutis, J. M. (2019). Self-reported face recognition is highly valid, but alone is not highly discriminative of prosopagnosia-level performance on objective assessments. *Behavior Research Methods, 51*(3), 1102-1116.

Aylward, E. H., Park, J. E., Field, K. M., Parsons, A. C., Richards, T. L., Cramer, S. C., & Meltzoff, A. N. (2005). Brain activation during face perception: evidence of a developmental change. *Journal of Cognitive Neuroscience*, *17*(2), 308–319.

Azevedo, R. (2005). Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning? The Role of Self-Regulated Learning. *Educational Psychologist*, *40*(4), 199–209.

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring Cognitive and Metacognitive Regulatory Processes During Hypermedia Learning: Issues and Challenges. *Educational Psychologist*, *45*(4), 210–223.

Bachmann, T., & Nurmoja, M. (2006). Are there affordances of suggestibility in facial appearance? *Journal of Nonverbal Behavior*.

Baron-Cohen, S. (1997a). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *The Maladapted Mind: Classic Readings in Evolutionary Psychopathology*, 207–239.

Baron-Cohen, S. (1997b). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 241–251.

Barragan-Jason, G., Lachat, F., & Barbeau, E. J. (2012). How Fast is Famous Face Recognition? *Frontiers in Psychology*, *3*, 454.

Banaji, M. R., & Dasgupta, N. (1998). The consciousness of social beliefs: A program of research on stereotyping and prejudice. In V. Y. Yzerbyt (Ed.), *Metacognition: Cognitive and social dimensions , (pp* (Vol. 253, pp. 157–170).

Bartlett, M., & Whitehill, J. (2010). Automated facial expression measurement: Recent applications to basic research in human behavior, learning, and education. *Handbook of Face Perception*.

Barzut, V., Markovic, S., & Zdravkovic, S. (2013). Face perception between race, gender and familiarity. Perception ECVP Abstract.

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211–241.

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006).

Eyewitness memory is still not common sense: comparing jurors, judges and law enforcement to eyewitness experts. Applied Cognitive Psychology, 20(1), 115–129.

Berry, D. S. (1991). Accuracy in social perception: contributions of facial and vocal information. *Journal of Personality and Social Psychology*, *61*(2), 298–307.

Berry, D. S., & Brownlow, S. (1989). Were the physiognomists right? Personality correlates of facial babyishness. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*.

Bindemann, M., Attard, J., & Johnston, R. A. (2014). Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychology*, *1*(1), 986903.

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96–103.

Bishop, S. L., & Seltzer, M. M. (2012). Self-reported autism symptoms in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*(11), 2354–2363.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2020). Computational insights into human perceptual expertise for familiar and unfamiliar face recognition. *Cognition*, 104341.

Bobak, A. K., Mileva, V. R., & Hancock, P. J. B. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. In *Quarterly Journal of Experimental Psychology* (Vol. 72, Issue 4, pp. 872–881).

Bonnen, K., Klare, B. F., & Jain, A. K. (2013). Component-Based Representation in Automated Face Recognition. IEEE Transactions on Information Forensics and Security, 8(1), 239–253.

Botwin, M. D., Buss, D. M., & Shackelford, T. K. (1997). Personality and mate preferences: five factors in mate selection and marital satisfaction. *Journal of Personality*, *65*(1), 107–136.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103*(4), 751–757.

Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: a critical review. *Cognitive Neuropsychology*, *17*(1), 55–71.

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20*(1), 24–27.

Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language*, *67*(1), 59–77.

Brown, A. L. (1977). Development, schooling and the acquisition of knowing about knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 241-253). Hillsdale, NJ: Lawrence Erlbaum A

Brown, A. L. (1978). Knowing when, where and, how to remember: A problem of metacognition. *Advances in Instructional Psychology*, *1*. http://ci.nii.ac.jp/naid/10012502747/

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*(2), 204–223.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*(1), 105–116.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R. & Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception, 22*(2), 131–152.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology. Applied*, *5*(4), 339.

Bruce, V., Henderson, Z., & Newman, C. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental, 7*(3), 207–18.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology , 77* (3), 305–327.

Bui, N. H. (2012). False consensus in attitudes toward celebrities. *Psychology of Popular Media Culture, 1*(4), 236–243.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22*(2), 153–176.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science. 23*(1), 1–31.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*(4), 943–958.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face Recognition in Poor-Quality Video: Evidence from Security Surveillance. *Psychological Science*, *10*(3), 243–248.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(4), 918–928.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, *40*(8), 1129–1138.

Calder, A., Rhodes, G., Johnson, M., & Haxby, J. (2011). *Oxford Handbook of Face Perception*. OUP Oxford.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–543.

Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A Critical Examination. *Child Development*, *53*(1), 11–28.

Cavazos, J. G., Phillips, P. J., Castillo, C. D., & O'Toole, A. J. (2020). Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *arXiv:1912.07398, 12 2019.*

Chellappa, R., Sinha, P., & Phillips, P. J. (2010). Face Recognition by Computers and Humans. *Computer, 43*(2), 46–55.

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. In *Visual Cognition* (Vol. 11, Issue 7, pp. 857–869).

Corrow, S. L., Davies-Thompson, J., Fletcher, K., Hills, C., Corrow, J. C., & Barton, J. J. S. (2019). Training face perception in developmental prosopagnosia through perceptual learning. *Neuropsychologia*, *134*, 107196.

Davies, S., Bishop, D., Manstead, A. S., & Tantam, D. (1994). Face perception in children with autism and Asperger's syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *35*(6), 1033–1057.

Dahl, C. D., Rasch, M. J., Tomonaga, M., & Adachi, I. (2013). Developmental processes in face perception. *Scientific Reports*, *3*, 1044.

Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience*, *10*, 10–20.

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology*, *32*(4), 331–341.

Dean, J. (2007). *Why we all stink as intuitive psychologists: The False Consensus Bias". PsyBlog*. Retrieved 2007-11-13.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, *28*(6), 687–706.

DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements in prosopagnosia: successes and failures over the last 50 years. *Frontiers in Human Neuroscience*, *8*, 561.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*(3), 987–1006.

De Heering, A., Rossion, B., Maurer, D. (2012). Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. *Cognitive Development,* 27, 17–27.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, *3*(1), 1–8.

Diaz, A. L. (2008). Do I know you? A case study of prosopagnosia (face blindness). *The Journal of School Nursing: The Official Publication of the National Association of School Nurses*, *24*(5), 284–289.

Dimaggio, G., Lysaker, P. H., Carcione, A., Nicolò, G., & Semerari, A. (2008). Know yourself and you shall know the other… to a certain extent: Multiple paths of influence of self-reflection on mindreading. *Consciousness and Cognition*, *17*(3), 778–789.

Dimaggio, G., Semerari, A., Falcone, M., Nicolo, G., Carcione, A., & Procacci, M. (2002). Metacognition, states of mind, cognitive biases, and interpersonal cycles: Proposal for an integrated narcissism model. *Journal of Psychotherapy Integration*, *12*(4), 421.

Duchaine, B., Jenkins, R., Germine, L., & Calder, A. J. (2009). Normal gaze discrimination and adaptation in seven prosopagnosics. *Neuropsychologia*, *47*(10), 2029–2036.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87.

Dunn, J. D., Summersby, S., Towler, A., Davis, J. P., & White, D. (2020). UNSW Face Test: A screening tool for super-recognizers. *PloS One*, *15*(11), e0241747.

Edens, K., & Potter, E. (2007). The Relationship of Drawing and Mathematical Problem Solving: Draw for Math Tasks. *Studies in Art Education*, *48*(3), 282–298.

Ekman, P., & Friesen, W. V. (1976). *Pictures of Facial Affect*. Consulting psychologists Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). CHAPTER II - How Do We Determine whether Judgments of Emotion Are Accurate? In P. Ekman, W. V. Friesen, & P. Ellsworth (Eds.), *Emotion in the Human Face* (Vol. 11, pp. 15–19). Pergamon.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*(4), 431–9.

Ellis, S., & Kruglanski, A. W. (1992). Self as an Epistemic Authority: Effects on Experiential and Instructional Learning. *Social Cognition*, *10*(4), 357–375.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is“ special” about face perception? Psychological Review. http://psycnet.apa.org/journals/rev/105/3/482/

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005).

Newborns' preference for face-relevant stimuli: effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(47), 17245–17250.

Feld, J., Sauermann, J., & de Grip, A. (2017). Estimating the relationship between skill and overconfidence. *Journal of Behavioral and Experimental Economics*, *68*, 18–24.

Fernbach, P., & Sloman, S. (2017). The knowledge illusion. Penguin.

Fine, D. R. (2012). A life with prosopagnosia. *Cognitive neuropsychology*, *29*(5-6), 354-359.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive--developmental inquiry. *The American Psychologist*, *34*(10), 906.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1280–1286.

Freeman, J. B., & Ambady, N. (2011). When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology*, *47*(1), 259–263.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495.

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1599), 2213–2223.

Frith, C. D., & Frith, U. (1999). Interacting minds--a biological basis. *Science*, *286*(5445), 1692–1695.

Fuocco, M. A. (1996, March 21). Trial and error: They had larceny in their hearts, but little in their heads. *Pittsburgh Post-Gazette*, p. Dl.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42.

Fussell, S. R., & Krauss, R. M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, *25*(3), 203–219.

Gage, N. M., & Baars, B. (2018). *Fundamentals of Cognitive Neuroscience: A Beginner's Guide*.

Ge, L., Luo, J., Nishimura, M., & Lee, K. (2003). The lasting impression of chairman Mao: hyperfidelity of familiar-face memory. *Perception*, *32*(5), 601–614.

Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A., Smoller, J. W., Kwok, H., Anthony, S. E., Nakayama, K., Rhodes, G., & Wilmer, J. B. (2015). Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology, 25*(20), 2684–2689.

Gignac, G. E., & Zajenkowski, M. (2020). The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. *Intelligence, 80*, 101449.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78*(2), 211–222.

Goodman, L. A., & Kruskal, W. H. (1979). Measures of Association for Cross Classifications. In L. A. Goodman & W. H. Kruskal (Eds.), *Measures of Association for Cross Classifications* (pp. 2–34). Springer New York.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549.

Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & David Smith, J. (2009). Metacognition is prior. *The Behavioral and Brain Sciences*, *32*(2), 142–142.

Grabman, J. H., Dobolyi, D. G., & Berelovich, N. L. (2019). Predicting high confidence errors in eyewitness memory: The role of face recognition ability, decision-time, and justifications. *Journal of Applied Research in Memory and Cognition*, 8, 233–243.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *The American Psychologist*, *35*(7), 603.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111.

Haan, E. H. F., Young, A., & Newcombe, F. (1987). Face recognition without awareness. *Cognitive Neuropsychology*.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (2009). *Handbook of Metacognition in Education*. Routledge.

Hadjikhani, N., Hoge, R., Snyder, J., & de Gelder, B. (2008). Pointing with the eyes: the role of gaze in communicating danger. *Brain and Cognition*, *68*(1), 1–8.

Hancock, P. J., Bruce, V., V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337.

Hancock, J. T., & Toma, C. L. (2009). Putting your best face forward: The accuracy of online dating photographs. *Journal of Communication, 59*(2), 367–386.

Hanley, J. R., Pearson, N. A., & Young, A. W. (1990). Impaired memory for new visual forms. *Brain: A Journal of Neurology*, *113 (Pt 4)*, 1131–1148.

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino (Ed.), *Handbook of motivation and cognition, Vol* (Vol. 3, pp. 28–84). The Guilford Press, xxvi.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216.

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, *11*(11), 2351–2355.

Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *The Journal of Consumer Research*, *21*(1), 205–218.

Haxby, J. V., & Gobbini, M. (2007). *The perception of emotion and social cues in faces*. https://psycnet.apa.org/record/2006-22475-001

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Hay, D. C., Young, A. W., & Ellis, A. W. (1991). Routes through the face recognition system. *The Quarterly Journal of Experimental Psychology Section A, 43*(4), 761–791.

Hayashi, H., & Nishikawa, M. (2019). Egocentric bias in emotional understanding of children and adults. *Journal of Experimental Child Psychology, 185*, 224–235.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689.

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology, 113*(4), 513–529.

Herlitz, A., & Yonker, J. E. (2002). Sex differences in episodic memory: the influence of intelligence. *Journal of Clinical and Experimental Neuropsychology*, *24*(1), 107–114.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *The Behavioral and Brain Sciences*, *21*(1), 101–114; discussion 115–148.

Hills, P. J. (2012). A developmental study of the own-age face recognition bias in children. *Developmental Psychology*, *48*(2), 499–508.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology. Applied*, *5*(2), 205–221.

Hole, G., & Bourne, V. (2010). *Face Processing: Psychological, Neuropsychological, and Applied Perspectives*. OUP Oxford.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84.

Hollingworth, H. L. (1922). *Judging human character*.

Hönekopp, J. (2006). Once more: is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 199–209.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions. *PloS One*, *10*(8), e0136905.

Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition*, *1*(3), 158–162.

Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgement of learning. *Memory* , *24*(3), 383–398.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*(4), 730–742.

Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I thought you were looking at me: direction-specific aftereffects in gaze perception. *Psychological Science*, *17*(6), 506–513.

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science, 319*(5862), 435–435.

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society B, 285*(1888), 20181319.

Jenkins, R., & Kerr, C. (2013). Identifiable images of bystanders extracted from corneal reflections. *PloS One*, *8*(12), e83325.

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116–119.

John, D., & Janet, M. (2009). *Metacognition*. Sage Publications, Amerika.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews. Neuroscience*, *6*(10), 766–774.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: a review. *Memory* , *17*(5), 577–596.

Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social Metacognition: An Expansionist Review. In *Personality and Social Psychology Review*, *2*(2), 137–154.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1476), 2109–2128.

Kaufmann, J. M., & Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception, 33*(4), 399–408.

Kelley, C. M., & Jacoby, L. L. (1996). Adult Egocentrism: Subjective Experience versus Analytic Bases for Judgment. *Journal of Memory and Language*, *35*(2), 157–175.

Kelly, K. J., & Metcalfe, J. (2011). Metacognition of emotional face recognition. *Emotion* , *11*(4), 896–906.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, *25*(1), 37–52.

Keysar B, Lin S, Barr DJ. (2003). Limits on theory of mind use in adults. *Cognition. 89*(1), 25-41.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & Cognition*, 12, 60–70.

Kleitman, S., & Stankov, L. (2001). Response selection strategies, metacognitive beliefs, and Confidence: An individual differences perspective. *Published Proceedings of the Inaugural International Conference: Self-Concept Theory, Research and Practice: Advances for the New Millennium, Sydney, Blue Mountains, 5--6 October, Australia*.

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: an ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67*(4), 596–610.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.

Kuhn, D. (1983). On the Dual Executive and Its Significance in the Development of Developmental Psychology. In *On the Development of Developmental Psychology* (Vol. 8, pp. 81–110). Karger Publishers.

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. *Children's Reasoning and the Mind*, 301–326.

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, *10*(3), 294–340.

Laidlaw, K. E. W., & Kingstone, A. (2017). Fixations to the eyes aids in facial encoding; covertly attending to the eyes does not. *Acta Psychologica*, *173*, 55–65.

Legrand, D., & Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, *116*(1), 252–282.

Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. *Journal of Comparative Psychology*. https://psycnet.apa.org/getdoi.cfm?doi=10.1037/a0019460

Lindsay, R. C. L., Mansour, J. K., Bertrand, M. I. Kalmet, N. & Melsom, E. I. (2011). Face recognition in eyewitness memory. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 307–328). Oxford University Press.

Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *Quarterly Journal of Experimental Psychology* , *71*(5), 1260–1262.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law, 7*(1), 98–118.

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy*, *10*(2), 3.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430.

Mason, S. E. (1986). Age and gender as factors in facial recognition and identification. *Experimental Aging Research*, *12*(3), 151–154.

Martinez, A. M. (2017). Computational Models of Face Perception. *Current Directions in Psychological Science*, *26*(3), 263–269.

Martinez, M. E. (2006). What is metacognition? *Phi Delta Kappan*, *87*(9), 696–699.

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. (2018). *Retrospective confidence judgments across tasks: domain-general processes underlying metacognitive accuracy.* https://doi.org/10.31234/osf.io/dr7ba

Metcalfe, J., & Dunlosky, J. (2009). *Metacognition: A textbook for cognitive, educational, life span & applied psychology*. Thousand Oaks, CA: Sage.

McKone, E., & Robbins, R. (2011). Are Faces Special? In Calder, A., Rhodes, G., Johnson, M., & Haxby, J. (Eds.). (2011). *Oxford handbook of face perception*. Oxford University Press. (pp. 149–176).

McIntosh, R. D., Fowler, E. A., Lyu, T., & Della Sala, S. (2019). Wise up: Clarifying the role of metacognition in the Dunning-Kruger effect. *Journal of Experimental Psychology. General*, *148*(11), 1882–1897.

McNeill, D. (1966). A study of word association. *Journal of Verbal Learning and Verbal Behavior*, 5, 548–557.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. Memory & Cognition, 34, 865–876.

Meichenbaum, D., & Asarnow, J. (1979). Cognitive-behavioral modification and metacognitive development: Implications for the classroom. *Cognitive-Behavioral Interventions: Theory, Research, and Procedures*, 11–35.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35.

Merluzzi, T. V., Rudy, T. E., & Glass, C. R. (1981). The information processing paradigm: Implications for clinical science. *Cognitive Assessment*, 77–124.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(4), 630–633.

Moon, H., & Phillips, P. J. (2001). Computational and performance aspects of PCA-based face-recognition algorithms. *Perception*, *30*(3), 303–321.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review*, *98*(2), 164–181.

Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social*

Science & Medicine, *211*, 274–281.

Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience, 12*, 303–310.

Nakashima, S. F., Langton, S. R. H., & Yoshikawa, S. (2012). The effect of facial expression and gaze direction on memory for unfamiliar faces. *Cognition & Emotion*, *26*(7), 1316–1325.

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology, 102*(4), 726–747.

Neil v. Biggers (1972). 409 U.S. 188.

Nelson, T. and Narens, L. (1990). Metamemory: a theoretical framework and new findings. *The psychology of learning and motivation*, 26:125–173.

Nelson, T. O., Kruglanski, A. W., & Jost, J. T. (1998). Knowing thyself and others: Progress in metacognitive social psychology. In V. Y. Yzerbyt (Ed.), *Metacognition: Cognitive and social dimensions , (pp* (Vol. 253, pp. 69–89).

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme, 17*(2), 141–161.

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220.

Nikolov, D., Oliveira, D. F., Flammini, A., & Menczer, F. (2015). Measuring online social bubbles. *PeerJ Computer Science, 1*, e38.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, 16, 389–408.

Njemanze, P. (2004). Transcranial doppler spectroscopy for assessment of brain cognitive functions (USPTO Patent No. 20040158155:A1). In *US Patent* (No. 20040158155:A1).

Noyes, E., & Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. *Cognition*, *165*, 97–104.

Noyes, E., & Jenkins, R. (2019). Deliberate disguise in face identification. *Journal of Experimental Psychology. Applied*, *25*(2), 280–290.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83–110.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092.

Oruc, I., Balas, B., & Landy, M. S. (2019). Face perception: A brief journey through recent discoveries and current directions. *Vision Research*, *157*, 1–9.

O'Toole, A. J. (2011). Cognitive and computational approaches to face recognition. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 15–30). Oxford, England: Oxford University Press.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22*(2), 208–224.

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., Irons, J., Al-Janabi, S., Taylor, L. C., Rivolta, D., & McKone, E. (2017). Do people have insight into their face recognition abilities? *Quarterly Journal of Experimental Psychology* , *70*(2), 218–233.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13.

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203.

Peirce, C. S., & Jastrow, J. (1884). *On small differences in sensation*. https://philarchive.org/archive/PEIOSD

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*(6), 1774–1784.

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a “kernel of truth” in social perception. *Social Cognition*, *24*(5), 607–640.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261.

Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of meta-cognition in social judgment. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 254– 284). New York, NY: Guilford.

Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., Banaji, M.R. (2000). Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation. *Journal of Cognitive Neuroscience*, 12, 729–738.

Phillips, P. J., Boyer, K. W., Flynn, P. J., O’Toole, A. J., Phillips, P. J., Schott, C. L., Scruggs, W. T., & Sharpe, M. (2007). *FRVT 2006 and ICE 2006 large-scale results*. Technical Report NISTIR 7408, NIST, 2007.

Phillips, P. J., Jiang, F., Narvekar, A., Ayyad, J., & O’Toole, A. J. (2011). An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception*, *8*(2), 1–11.

Phillips, P. J., Jonathon Phillips, P., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J.-C., Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. In *Proceedings of the National Academy of Sciences* (Vol. 115, Issue 24, pp. 6171–6176).

Phillips, P. J., Yates, A. N , Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J, Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences, 115*(24), 6171–6176.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *209*(4), 481–493.

Pound, N., Penton-Voak, I. S., & Brown, W. M. (2007). Facial symmetry is positively associated with self-reported extraversion. *Personality and Individual Differences*, *43*(6), 1572–1582.

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences.11*(1), 37–43.

Ramon, M., Caharel, S., & Rossion, B. (2011). The speed of recognition of personally familiar faces. *Perception*, *40*(4), 437–449.

Ramon, M., Sokhn, N., & Caldara, R. (2019). Decisional space modulates visual categorization–Evidence from saccadic reaction times. *Cognition, 186*, 42–49.

Ranjan, R., Sankaranarayanan, S., Bansal, A., Bodla, N., Chen, J., Patel, V. M., Castillo, C. D., & Chellappa, R. (2018). Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans. *IEEE Signal Processing Magazine*, *35*(1), 66–83.

Rausch, M., Müller, H. J., & Zehetleitner, M. (2015). Metacognitive sensitivity of subjective reports of decisional confidence and visual experience. *Consciousness and Cognition*, *35*, 192–205.

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174.

Rhodes, M. G., Sitzman, D. M., & Rowland, C. A. (2013). Monitoring and control of learning own-race and other-race faces. *Applied Cognitive Psychology*, *27*(5), 553–563.

Riddoch, M. J., Johnston, R. A., Martyn Bracewell, R., Boutsen, L., & Humphreys, G. W. (2008). Are faces special? A case of pure prosopagnosia. In *Cognitive Neuropsychology* (Vol. 25, Issue 1, pp. 3–26).

Ritchie, K. L., Kramer, R. S., & Burton, A. M. (2018). What makes a face photo a ‘good likeness’? *Cognition, 170*, 1–8.

Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition*, *141*, 161–169.

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face Recognition by Metropolitan Police Super-Recognisers. *PloS One*, *11*(2), e0150036.

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *The British Journal of Educational Psychology*, *79*(4), 749–767.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274–289.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*(6), 440–444.

Saraç, S., & Karakelle, S. (2017). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education*, *4*(2), 301–315.

Sauerland, M., Sagana, A., Siegmann, K., Heiligers, D., Merckelbach, H., & Jenkins, R. (2016). These two are different. Yes, they're the same: Choice blindness for facial identity. *Consciousness and Cognition*, *40*, 93–104.

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution, and other dispatches from the wrongly convicted.* New York: Random House.

Schellings, G. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning*, *6*(2), 91–109.

Schnitzspahn, K. M., Zeintl, M., Jäger, T., & Kliegel, M. (2011). Metacognition in prospective memory: are performance predictions accurate? *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *65*(1), 19–26.

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475.

Schriber, R. A., Robins, R. W., & Solomon, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. *Journal of Personality and Social Psychology*, *106*(1), 112–130.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience, 23*(2-3), 125–141.

Schwartz, B. L. (2002). The phenomenology of naturally-occurring tip-of-the-tongue states: A diary study. *Advances in Psychology Research*, *8*, 71–84.

Scopelliti, I., Morewedge, C. K., McCormick, E., Min, H. L., Lebrecht, S., & Kassam, K. S. (2015). Bias Blind Spot: Structure, Measurement, and Consequences. *Management Science*, *61*(10), 2468–2486.

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science, 2*(6), 140343.

Shapiro, S. S. (1966). Word association norms: Stability of response and chains of association. *Psychonomic Science*, *4*(6), 233–234.

Shriver, E.R., Young, S.G., Hugenberg, K., Bernstein, M.J., Lanter, J.R. (2008). Class, Race, and the Face: Social Context Modulates the Cross-Race Effect in Face Recognition. *Personality and Social Psychology Bulletin*, 34, 260–274.

Simons, D. J. (2013). Unskilled and optimistic: overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin & Review*, *20*(3), 601–607.

Simion, F., & Giorgio, E. D. (2015). Face perception and processing in early infancy: inborn predispositions and developmental changes. *Frontiers in Psychology*, *6*, 969.

Simpson, W. E., & Crandall, S. J. (1972). The perception of smiles. *Psychonomic Science*, *29*(4), 197–200.

Smith, S. M., Stinson, V., & Prosser, M. A. (2004). Do they all look alike? An exploration of decision-making strategies in cross-race facial identifications. *Canadian Journal of*

Behavioural Science/Revue Canadienne Des Sciences Du Comportement, *36*(2), 146.

Smith, A. D., Hood, B. M., & Hector, K. (2006). Eye remember you two: gaze direction modulates face recognition in a developmental study. *Developmental Science*, *9*(5), 465–472.

Soh, L., & Jacobs, K. E. (2013). The biasing effect of personality on self-estimates of cognitive abilities in males and females. *Personality and Individual Differences*, *55*(2), 141–146.

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and Self-Regulated Learning Constructs. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *10*(2), 117–139.

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. *The Oxford Handbook of Thinking and Reasoning*, 343–365.

Stanovich, K. E., & West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? In *Heuristics and Biases* (pp. 421–440). https://doi.org/10.1017/cbo9780511808098.026

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology. General*, *127*(2), 161.

Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research.

Current Opinion in Neurobiology, *23*(3), 423–429.

Sutherland, C. A., Burton, N. S., Wilmer, J. B., Blokland, G. A., Germine, L., Palermo, R., Collova, J. R., & Rhodes, G. (2020). Individual differences in trust evaluations are shaped mostly by environments, not genes. *Proceedings of the National Academy of Sciences, 117*(19), 10218–10224.

Sutherland, C. A., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*(4), 521–537.

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118.

Sutherland, C. A., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality? *British Journal of Psychology, 111*(2), 215–232.

Tanaka, K. 'ichiro. (1993). Egocentric bias in perceived fairness: Is it observed in Japan? *Social Justice Research*, *6*(3), 273–285.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *46*(2), 225–245.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*(1), B1–B9.

Tang, X., Wang, X., 2004. Face Sketch Recognition. *IEEE Transactions on Circuits and Systems for Video Technology,* 14, 50–57.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*(4), 1115–1143.

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: evidence from the face inversion effect. *Vision Research*, *51*(11), 1273–1278.

Tenenberg, J., & Murphy, L. (2005). Knowing what I know: An investigation of undergraduate knowledge and self-knowledge of data structures. *Computer Science Education*, *15*(4), 297–315.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Tobias, S., & Everson, H. (1996). Assessing Metacognitive Knowledge Monitoring. Report No. 96-01. *College Entrance Examination Board*. https://eric.ed.gov/?id=ED562584

Todorov, A. T., Said, C. C., & Verosky, S. C. (2011). Personality Impressions from Facial Appearance. In G. Rhodes, A. Calder, M. Johnson, & J. V. Haxby (Eds.), *Oxford*

Handbook of Face Perception. Oxford University Press.

Tokuhama-Espinosa, T. (2014). *Making Classrooms Better: 50 Practical Applications of Mind, Brain, and Education Science*. W. W. Norton & Company.

Towler, A., Kemp, R. I., Burton, A. M., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work? *PloS One*, *14*(2), e0211037.

Towler, A., Kemp, R. I., & White, D. (2017). Unfamiliar face matching systems in applied settings. In *Face Processing: Systems, Disorders and Cultural Difference,*. Bindemann, M. & Megreya, A.M. (Eds.), Nova Science.

Towler, A., White, D., & Kemp, R. I. (2014). Evaluating training methods for facial image comparison: The face shape strategy does not work. *Perception, 43*(2-3), 214–218.

Towler, A., White, D., & Kemp, R. (2020). *Can face identification ability be trained? Evidence for two routes to expertise*. . https://doi.org/10.31234/osf.io/g7qfd

Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, *95*, 124–137.

Tüttenberg, S. C., & Wiese, H. (2019). Learning own- and other-race facial identities from natural variability. *Quarterly Journal of Experimental Psychology* , *72*(12), 2788–2800.

Ueda, S., & Koyama, T. (2011). Influence of eye make-up on the perception of gaze direction. *International Journal of Cosmetic Science*, *33*(6), 514–518.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43*(2), 161–204.

Van Essen, G., & Hamaker, C. (1990). Using Self-Generated Drawings to Solve Arithmetic Word Problems. *The Journal of Educational Research*, *83*(6), 301–312.

Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice: A Publication of the Division for Learning Disabilities, Council for Exceptional Children*, *18*(4), 246–254.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences*, *15*(2), 159–176.

Vickers, D. (2014). *Decision Processes in Visual Perception*. Academic Press.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness Evidence: Improving Its Probative Value. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *7*(2), 45–75.

Wheeler, F. W., Liu, X., & Tu, P. H. (2011). Face recognition at a distance. *Handbook of Face Recognition*. https://link.springer.com/chapter/10.1007/978-0-85729-932-1_14

Wheeler, F. W., Weiss, R. L., & Tu, P. H. (2010). Face recognition at a distance system for surveillance applications. *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 1–8.

White, D., Burton, A. L., & Kemp, R. I. (2016). Not looking yourself: The cost of self-selecting photographs for identity verification. *British Journal of Psychology, 107*(2), 359–373.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error Rates in Users of Automatic Face Recognition Software. *PloS One*, *10*(10), e0139827.

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100–106.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS One*, *9*(8), e103510.

White, D., Sutherland, C. A., & Burton, A. L. (2017). Choosing face: The curse of self in profile image selection. *Cognitive research: principles and implications, 2*(1), 1–9.

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. *Psychological Science, 30*(2), 261–272.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238-5241.

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science*, *26*(3), 225-230.

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*(3), 1830–1839.

Wirth, B. E., & Carbon, C.-C. (2017). An easy game for frauds? Effects of professional experience and time pressure on passport-matching performance. *Journal of Experimental Psychology. Applied*, *23*(2), 138–157.

Wolfson, S. (2000). Students' estimates of the prevalence of drug use: Evidence for a false consensus effect. *Psychology of Addictive Behaviors, 14*(3), 295–298.

Wright, P. (2002). Marketplace Metacognition and Social Intelligence. *Journal of Consumer Research, 28*(4), 677–682.

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: a problem of recognition. *Journal of Psychosomatic Research*, *65*(5), 445–451.

Yarmey, A. D. (1971). Recognition memory for familiar “public” faces: Effects of orientation and delay. Psychonomic Science, 24, 286–288.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1310–1321.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141.

Yinon, Y., Mayraz, A., & Fox, S. (1994). Age and the false-consensus effect. *The Journal of Social Psychology, 134*(6), 717–725.

Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science. 26*(3), 212–217.

Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury. In *Brain* (Vol. 116, Issue 4, pp. 941–959).

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and Voice Perception: Understanding Commonalities and Differences. *Trends in Cognitive Sciences*, *24*(5), 398–410.

Young, A. W., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). Facial expressions of emotion: Stimuli and tests (FEEST). *Bury St. Edmunds: Thames Valley Test Company*.

Zaki, S. R., & Johnson, S. A. (2013). The role of gaze direction in face memory in autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *6*(4), 280–287.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, *2*(3), 1497.

Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. *Cognition, 203*, 104345.

Zhou, X., Mondloch, C.J. (2016). Recognizing “Bella Swan” and “Hermione Granger”: No Own-Race Advantage in Recognizing Photos of Famous Faces. *Perception*, 45, 1426–1429.